Blame the Data or Blame the Theory?

On (In-) Comparability in International Survey Research

Insa Bechert

Blame the Data or Blame the Theory?

On (In-) Comparability in International Survey Research

Insa Bechert

Inaugural dissertation submitted in partial fulfillment of the requirements

for the degree Doctor of Social Sciences

at the University of Mannheim

Mannheim, 2018

Dean of the School of Social Sciences:	Prof. Dr. Michael Diehl
Advisor of the Dissertation:	Prof. Dr. Christof Wolf
Co-Advisor of the Dissertation:	Prof. Dr. Annette Schnabel
Reviewers of the Dissertation:	Prof. Dr. Michael Braun
	Prof. Dr. Annette Schnabel
Committee of the Doctoral Defense:	Prof. Dr. Michael Braun
	Prof. Dr. Christof Wolf
	Prof. Dr. Annette Schnabel
Date of Defense:	March 22nd, 2018

Contents

Acknowledgements **5**

1. Frame **7**

1.1 Theoretical Framework and Introduction 7

1.2 Connecting my Thesis to the Theoretical Framework 11

1.3 Literature Review 14

1.4 Extended Abstracts of the Three Articles 18

1.5 Conclusion 24

1.6 References 25

2. Article I: ***published***

Observing Unexpected Patterns in Cross-National Research: Blame Data, Theory, or Both?
Attitudes toward Redistributive Taxation in Thirty-Three Countries

Abstract 31

2.1 Introduction 32

2.2 Socioeconomic Status and Attitudes toward Redistributive Taxation: Why should Relationships Differ Systematically across Countries? 33

2.2.1 A Western Bias in Cross-National Research? 33

2.2.2 Explaining Cross-National Variation: The Roles of the Welfare State and Economic Development 34

2.3 Data, Methods, and Variables 37

2.4 Results 40

2.4.1 Attitudes toward Taxation in Thirty-Three Countries 40

2.4.2 Relationships between Socioeconomic Status and Attitudes toward Taxation in Thirty-Three Countries 43

2.4.3 Venezuela, South Africa, and the Philippines: Measurement Errors Related to Translation or Sample Coverage? 47

2.5 Conclusion 50

2.6 References 53

Appendix 58

Statement of Co-Author 61

3. Article II: *published*

Comparing Religiosity Cross-Nationally - *About Invariance and the Role of Denomination*

Abstract ... 63

3.1 Introduction ... 64

3.2 Profiles of Religiosity ... 65

3.3 Religion and Denomination ... 67

3.4 Hypotheses ... 68

3.5 Data, Variable Selection, and Methods ... 70

3.5.1 Data Source ... 70

3.5.2 Variables ... 70

3.5.3 Latent Class Analysis ... 73

3.6 Results ... 74

3.6.1 Results I: 35-country LCA: Profile Description and Invariance Test ... 74

3.6.2 Results II: Finding a Structurally Homogeneous Subgroup ... 76

3.6.3 Results III: Distribution of Denominations across Profiles ... 80

3.7 Discussion… ... 82

3.7.1 …on Comparability ... 82

3.7.2…on the Impact of Denomination ... 82

3.8 Conclusion and Outlook ... 84

3.9 References ... 85

3.9.1 Data ... 85

3.9.2 Literature ... 85

Appendix I ... 89

Appendix II ... 91

Appendix III ... 92

4. Article: *under review*

Is there a Continuous Way to God? *A Cross-National Evaluation of the Indicator "Self-Assessed level of Religiosity" as a Continuous Single-Item Measure of Religiosity*

Abstract ... 93

4.1 Introduction ... 94

4.2 Measures ... 96

4.2.1 Single-Item vs. Multi-Item Measures ... 96

4.2.2 Categorical vs. Continuous Measures ... 97

4.2.3 Self-Assessment Measures ... 99

4.3 Research Questions and Hypotheses 100
4.4 Data, Country Choice, Variables, and Methods 101
4.4.1 Data and Countries 101
4.4.2 SAR 101
4.4.3 Variables for the Multi-Item Measures 101
4.4.4 Dependent Variables for the Regressions 102
4.4.5 The Multi-Item Measures 103
4.5 Results 104
4.5.1 Step One: Correlation, Variance Analyses, and Scheffe Test Results 105
4.5.2 Step Two: Regression Results 107
4.6 Discussion 109
4.7 Conclusion 112
4.8. References 114
4.8.1 Data 114
4.8.2 Literature 114
Appendix 118

Acknowledgements

The starting point for this dissertation project was a conversation I had with Knut Kalgraff Skjåk during a coffee break of an ISSP meeting about the initial ideas of my first article. Knut gave me the opportunity to present these ideas at the 2009 ESRA conference held in Warsaw. After that, it was a long way to finishing this dissertation and it would not have succeeded without the encouragement, support and intellectual input of many people I am privileged to know:

First, I would like to thank my research supervisor Christof Wolf for his interest in my work and the time he invested. Most of all I truly appreciate that I could always count on his support and guidance throughout the developing process of my doctoral research project. I would also like to thank my second supervisor Annette Schnabel for her encouragement, time and advice especially during the last phase of my dissertation.

My sincere thanks also to Markus Quandt for granting me the flexibility required to accomplish my dissertation in addition to my routine work obligations. But I am even more grateful for the constructive discussions we had and valuable advice he gave me on many occasions. Special thanks go to Jonas Edlund who enriched my methodological knowledge a great deal and has always been an inspiration given his natural personal curiosity and excitement for survey data analyses.

I have great colleagues at the GESIS Data Archive for the Social Sciences and I am most grateful for the feedback given when I presented my work at the department's colloquium. There was always an open door and the willingness to discuss whatever research problem was troubling me. I am particularly grateful to Irene Müller and Petra Brien for their understanding when my priorities were on the thesis and not on my regular work, especially during the past year. Thank you as well to Urs Fichtner and Christoph Schulte who were there with helping hands, for so many little, yet so important, things in the process.

I should probably also thank several anonymous reviewers for their comments on my papers. Though at first an annoyance, albeit a necessary one, in the end, most of them led to improvements and eventually publications.

There is also the mental support and encouragement given by Doris Richter gen. Kemmermann and Kristi Winters for which I am most grateful. They were a source of stability when things did not proceed as expected or at the speed I wanted.

Most importantly I thank my family for their support. My daughter Maya who showed as much patience and understanding as lies within the limits of her possibilities and my husband Jan for never allowing any doubts that I would finish this, I am grateful to you both. To my parents Gerda and Walter Kümmel, my eternal thanks for always standing behind me; it is your unwavering support which made this dissertation project possible in the first place. Thank you.

1. Frame

1.1 Theoretical Framework and Introduction

Over the last thirty years there has been a substantial increase in the number of cross-national surveys, along with a concomitant rise in the number of participating countries worldwide (an overview of international surveys can be found in GESIS (2017); their historical development has been described by Smith and Fu (2016)). These data provide researchers the opportunity to compare attitudes, opinions, values, and behavior across a diverse range of topics and a great variety of countries. Ideally, this results in establishing awareness, and supports the understanding of national and global realities. Naturally, among the increasing number of countries some of the countries are very different from one another. For social surveys, this means that theoretical ideas, translated into measurement instruments have to be applied to very different settings, i.e., countries composed of very different cultures, languages or historical backgrounds. Also to be considered are different basic technical prerequisites, such as the financial resources available for staff training, the characteristics of population registers from which a sample can be drawn, or ultimately even the distribution of communication devices used for interviews which may produce structural differences. Thus, the potential risk that the resulting data are not comparable across these settings and prerequisites is ubiquitous in international comparative research.

Almost 20 years ago Sir Roger Jowell, co-founder of the international comparative survey programs *European Social Survey* (ESS) and the *International Social Survey Programme* (ISSP), wrote an article titled *How Comparative is Comparative Research?* (Jowell 1998), aimed squarely at "underlining the pitfalls of cross-national measures, to urge greater caution in their interpretation and to suggest a number of partial remedies" (Jowell 1998, 168). My dissertation project has similar goals. It consists of three articles, all dealing with particular examples of cross-national (in)comparability of survey instruments. The "remedies" resulting from my research may be regarded as contributions to the overarching goal of ensuring greater caution in international comparative research. All articles are based on data taken from the ISSP. Between 1984 and 2017 the ISSP expanded from 5 to 48 countries across six continents and thus offers a significant amount of cultural dissimilarities for analyses. The ISSP is well coordinated and implements numerous measures[1] to ensure a high level of data quality and comparability. The effectiveness of these measures is indirectly reflected in my research and the resulting articles.

[1] For more information on the ISSP organizational structure and working principles see https://www.gesis.org/issp/home/.

Generally, when conducting surveys and especially when doing so in different countries to be compared later, numerous challenges and potential sources of error arise throughout the process (for a comprehensive discussion on these challenges see Johnson and Braun 2016). The potential error sources can be compiled and presented in numerous ways. The framework I am going to use for the systematic placement of the comparability aspects discussed in my dissertation articles is the *Total Survey Error* (TSE) approach which provides a coherent layout of the various types of potential survey error and sets a methodological research agenda for studying error structures in surveys (Smith 2017). Different TSE schemes have been developed to classify, describe and connect all errors that may arise in the course of a survey life cycle (see for example Groves 1989; Weisberg 2005; Groves et al. 2009; and Biemer et al. 2017). They cover potential survey errors[2] concerning the representation of the target group, as well as measurement related issues. Survey errors can lead to systematic bias and variance (Biemer 2010), both of which interfere with cross-national comparability.

The central idea of the TSE, however, focuses on stand-alone (national) and not on comparative surveys. Weisberg's (2005) paradigm extends this scope by including *comparison effects*. But, classified as *survey administration issues*, in their interpretation he focuses on the technical aspects of comparability challenges, such as house effects, arising via different survey organizations' use of different styles and strategies in the data collection process. Weisberg's scheme does not deal with the main challenge of comparative surveys: the comparison of different true values cross-nationally. In an effort to cover this, Smith (2011) proposes extending the TSE through the broad concept of *comparison error* which, by definition, takes into account the impact of intersocietal differences of language, culture and structure. It also includes comparability aspects specifically related to conducting the surveys, such as privacy laws, social desirability norms, civil liberties and survey climate on the different components of the survey lifecycle (Smith 2011, 475). The larger the number of countries, the higher the risk for comparison error (Smith 2017).

Also, Pennell et al. (2017) have discovered a shortcoming of the TSE paradigm regarding comparative surveys. They extend Smith's approach of adding a *comparison error* by explicitly developing a TSE paradigm for multinational, multiregional, and multicultural (3MC) contexts (Pennell et al. 2017, 182). This paradigm is based on Groves' et al. (2009) twofold survey lifecycle distinguishing *representation* of the target population and *measurement* of the construct of interest and lists the different potential sources of error that may contribute to comparison error across all components of 3MC surveys:

[2] *Error* should not be understood as a synonym for *mistake* here, but as a reference to the difference between obtained values and true values (Weisberg 2005, 18).

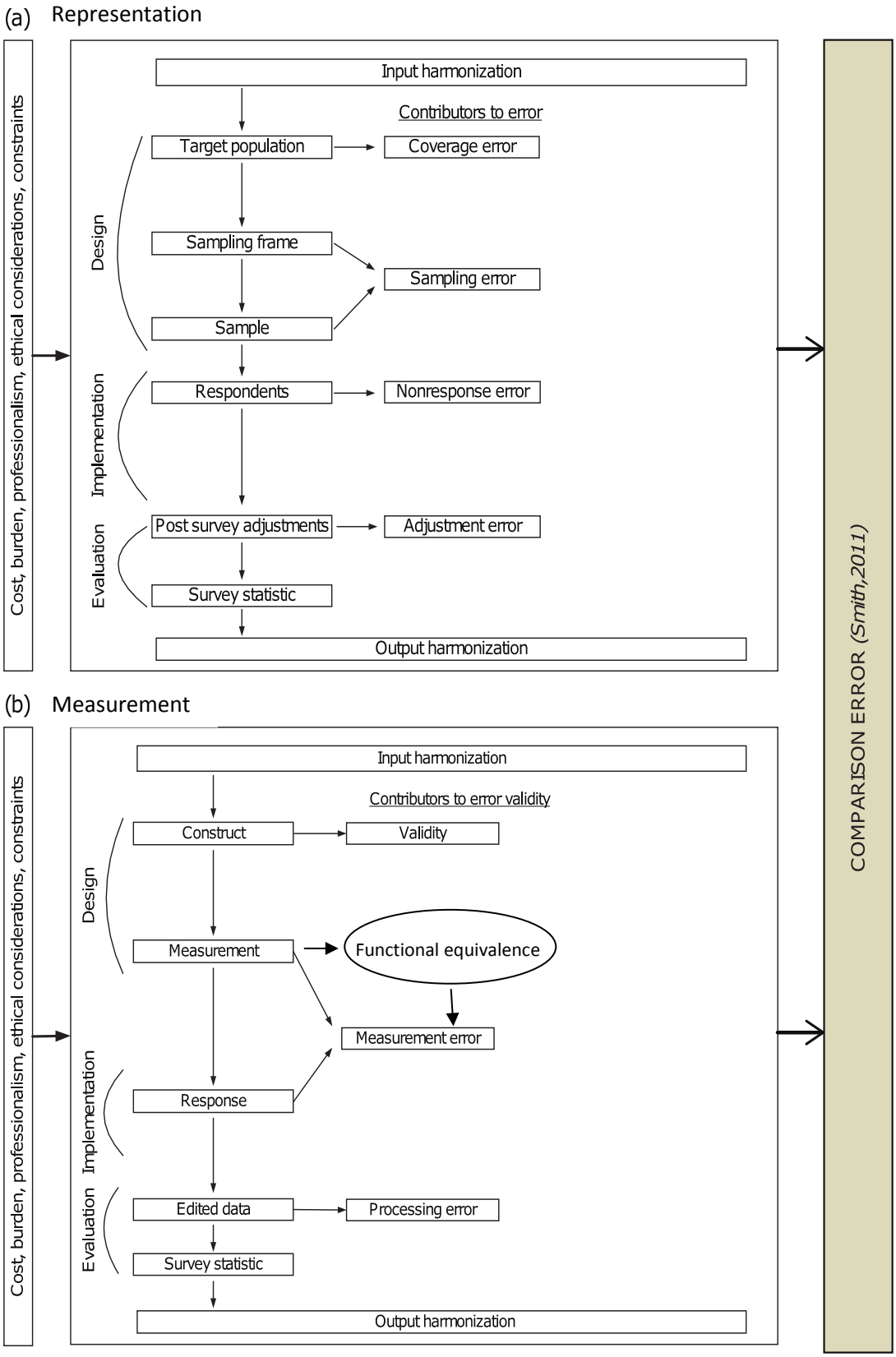

Figure 1: TSE (a) Representation and (b) Measurement in a cross-cultural context. Slightly modified figure by Pennell et al. 2017 adapted from Groves et al. (2009).

The TSE diagram shows impressively how large this field of potential error sources in survey research is with respect to comparison error. The *representation* and the *measurement* section

are both structured to have a design-, an implementation-, and an evaluation phase. The accurate *representation* of the target group may be compromised by coverage and sampling error in the design phase of the survey. In the implementation phase, numerous factors may cause unit non-response, and the main risk in the evaluation phase, leading to biased data, comes from incorrectly applied weighting procedures.

The main potential pitfalls in the *measurement* section's design phase of international surveys are the national- and cross-national validity of concepts and their adequate translation into functional equivalent measurement instruments. Functional equivalence is a key issue for comparability in cross-national surveys (Weisberg 2005, 298; Smith 2011, 475). Often the same measurement instrument cannot be utilized across countries because the phenomenon to be measured does not exist in the same way according to cultural or infrastructural differences. A prominent example is the question of just how literally the words of the Bible should be taken. This question has different implications depending on whether Christians or non-Christians are asked. In such cases, functional equivalent measurement instruments need to be found. Asking the respondents about the Holy Scriptures of their own religion is a possible solution for the Bible example. The usage of functional equivalent measures ensures that data can be compared across groups because the resulting measures reflect the same phenomenon (Hoffmeyer-Zlotnik and Wolf 2003, 389).

The *measurement* section's implementation phase focuses on the response process and all possible related sources of error. Nonetheless, once the data has been collected, all manner of processing errors in the evaluation phase could still compromise the data.

The TSE scheme by Pennell et al. also incorporates two means of harmonization: Input and Output harmonization, which play a unique role in comparative surveys, because the procedures directly aim at achieving or at least improving comparability (Pennell et al. 2017, 183). Both procedures can be useful, depending on the particular situation. However, input harmonization has the advantage of less error potential compared to output harmonization. If in the data collection process input harmonization is possible, survey programs will therefore prefer that. Because of the previously mentioned specific, historically grown survey structure, ISSP background variables to a certain degree depend on (ex-ante) output harmonization. *Costs, burden, professionalism, ethical considerations* and other design *constraints* may all have an impact on TSE elements and all elements could potentially add to *comparison error*.

It is beyond the scope of a single dissertation project to examine each potential sources of survey error using relevant examples. My contributions mainly address the design, and the implementation phases of the TSE measurement section (Figure 1b). More specifically, the

central focus of my articles is on measurement error and construct validity. After providing the theoretical outline for my thesis I will give an overview of the most prominent literature concerned with this particular field of survey errors due to measurement (Figure 1b) and their effects. This literature review is provided in an effort to shed some light on the scope of research in this specific area and to embed the elements addressed in my articles. The following section contains detailed abstracts of the individual articles including summaries of their contributions to handling, understanding and interpreting comparative data, and some ideas on connecting research steps. The final section concludes this framepaper with some recommendations regarding the awareness and handling of potential pitfalls encountered in international data comparisons.

1.2 Connecting my Thesis to the Theoretical Framework

A theory, plus two or more constructs or concepts[3], and appropriate measures are essential to examine a phenomenon. Figure 2 and 3 depict the levels on which different error sources can be located in the process. The theory level (Figure 2) is the point of departure. From the theory constructs are derived. They serve as links by "bringing theories down to earth" (Laerd 2016), or in this case to the construct/concept level, and helping to explain the theories' components. At the bottom, the observation level (Figure 3) is the basic level, where the single indicators for the measurement of constructs are located.

Following this hierarchy, the first question to be asked is whether constructs (X and Y) can validly be derived from theory and whether these constructs are equivalent across countries (Figure 2). If it is not possible to derive the same constructs in countries A, B and C from theory, the validity range of the theory has to be regarded as limited and for an unsuccessful research model, essentially, "the theory is to blame". As an example, let us assume an analysis includes attitudes towards same sex marriage. If same sex marriage is not legalized by some governments and therefore does not exist in these countries, responses may not be comparable.[4]

[3] In the literature there is disagreement regarding the differentiation between "construct" and "concept". Often both terms are used interchangeably (Billiet 2016: 194). Some authors assign higher levels of theoretical abstraction to the one, others to the other term. Since the differentiation is irrelevant in my context, I will refrain from a more detailed definition and in the following will refer to constructs.

[4] Of course, this is a very obvious example. The more veiled the differences, the more difficult they are to detect.

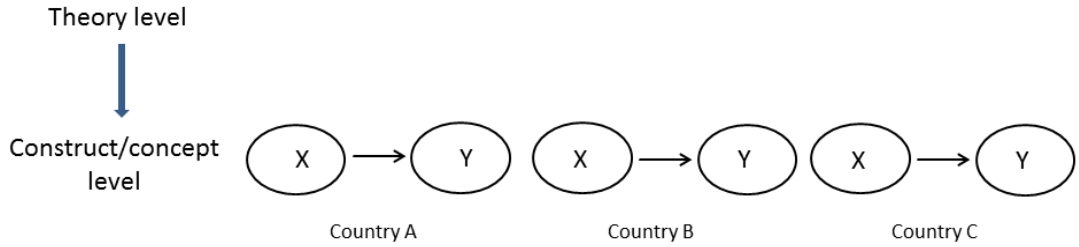

Figure 2: Blame the theory? Existence of concepts (X, Y) derived from theory and comparability across countries A, B and C

If the same constructs can be derived the questions remain whether the indicators used for operationalization of the construct are functionally equivalent across countries and whether the relationship between the indicator(s) and the construct is the same across countries. Both criteria must be met to fulfill the prerequisites for comparability. If they are not met, "the data is to be blamed" for incomparability.

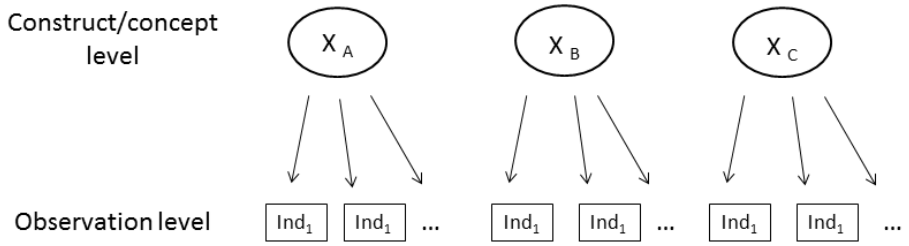

Figure 3: Blame the data? The similarity of the relationships between indicators and constructs across countries A, B and C

In my dissertation project I am seeking out sources of incomparability within the TSE framework on all three hierarchical levels, which are all prone to error, but entailing different consequences. However, the three articles of my dissertation are based on practice-oriented empirical research. For the theoretical framework this means that the single elements of the TSE paradigm are not necessarily treated in a systematical order. In the following, I will shortly outline how my analyses across the three articles take up TSE elements on the three hierarchical levels of a measurement model. The particular methodological approaches will be explained in more detail in the extended abstracts (section 1.4) of this frame paper and, of course, in the method sections of the articles.

One of the main challenges encountered in international research arises from the fact that many theories were initially developed for use solely in Western European and Anglo-American contexts without regard to their transposability to other cultures. Thus, an important research question discussed in my thesis is whether theories developed in a Western context have the same predictive power in non-Western contexts (Figure 2).

If research confirms a hypothesis, there is good reason to believe in the validity of the theory, the concept, and the measure. If it is not confirmed there could be any number of reasons. One way to narrow down the sources of error is by employing the principle of exclusion. It is a cumbersome business, requiring an individual and appropriate treatment of each potential error source, while taking into account that some error sources are more likely than others and some are easier to detect than others. In some cases, such a procedure is worth the effort since it provides a clear and comprehensive picture of both, data strengths and weaknesses. I use this approach to narrow down possible error sources by looking at a representation issue of the survey, in the form of coverage error. Additionally, I look for translation problems, a classical element of *measurement error*. Both error sources directly contribute to comparison error, if (1) the sample coverage in a country is not just incomplete, but in its incompleteness also differs from the sample coverage in the other countries; and (2) the translated question not only deviates from the defined standard, but in the process also creates a deviation from other countries (Figure 3).

By dealing with the question whether it is possible to validly compare the latent construct of religiosity cross-nationally, another analysis in my thesis, again, examines an error that deals with *measurement* directly related to *comparison error*. Comparisons do not necessarily have to be conducted across nations, but just as well across cultural, ethnic, denominational, or whatever groups are crucial for the subject of interest. I examine how indicators of religiosity are related to the latent construct of religiosity across countries and across denominations nested within countries (Figure 3).

Another examination looks at the validity and predictive potential of a specific indicator. If the indicator is not a suitable measurement instrument, because it turns out not to capture the initial concept, the error has to be classified as *measurement error* due to the operationalization of the construct. When the measurement instrument works fine but turns out not to be functionally equivalent across countries (Figure 3), the specific measurement, again, provokes *comparison error*.

When analyzing data for comparability the full range of the TSE is always kept in mind, because it can never be excluded that there are combined effects of different error sources (Smith 2011, 471). However, being aware of sources potentially causing incomparability is not enough. To examine the degree to which comparability across country data could actually be achieved, I test them for invariance (Figure 3).

1.3 Literature Review

A great deal of research has been done on theoretically exploring the whole spectrum of survey errors, empirically examining their effects and offering solutions to reduce them. All aspects are relevant for cross-national survey data quality because, taken together, they all contribute to compromising comparability. This body of literature, however, is much too large to present an encompassing overview here and so I will narrow down the field. Save for a quick side excursion to the *representation* section of survey error, by looking for sample coverage, my articles focus on the *measurement in cross-cultural context* section (Figure 1b). To stay focused, I will also restrict the following literature review on this part of the TSE framework, following its structure from top to bottom. Pertinent sources of reference covering the whole spectrum of TSE aspects are volumes on survey methodology, such as Wolf et al. (2016), Harkness et al. (2010), or the Cross-Cultural Survey Guidelines (Survey Research Center 2016).

Measurement in cross-cultural context

The discussion of potential measurement error sources has such a broad scope because measurement provides the required detour necessary in order to capture a concept. Or, as Jaak Billiet (2016) expresses it: "[…], the measure of a concept is not the concept itself, but one of several possible error-filled ways of measuring it" (Billiet 2016, 193). A construct can be defined as a more systematically and formally defined concept (Hox 1997, 48f.; Kerlinger 1973, 29; Billiet 2016, 195). The latent constructs situated in the focus of my work are attitudes towards redistribution and religiosity. Others have investigated the cross-national comparability of measurement instruments for other constructs such as education (Schneider 2009; Hoffmeyer-Zlotnik and Warner 2014; Ortmanns and Schneider 2016), labor status, household composition, (Hoffmeyer-Zlotnik and Warner 2011) or ethnic background (Aspinall 2007; Hoffmeyer-Zlotnik and Warner 2014; Heath et al. 2016) to mention but a few.

Validity

One key question for cross-national surveys is whether the concept of interest exists across countries. Barnes and Kaase (1979) raise the question of whether the phenomenon of mass participation in political action across five Western democracies really can be answered with only one theory. Authors, such as Drake (1973) or Wallerstein (1997) go one step further by pointing out the problem that quite often perspectives and approaches on scientific concepts and methodology from Western origin and contexts are applied to non-Western societies and data without reflection. More recently, Blasius and Thiessen formulated a cultural distance

hypothesis that goes in the same direction. It states that the greater the cultural distance between the origin of a survey instrument (based on a concept) and the group being investigated, the more compromised the data quality (Blasius and Theissen 2012, 5). Quite simply, if the concept does not exist in a country the measurement instrument cannot produce valid data.

To achieve validity, the concept also needs to be adequately operationalized. The development from concepts to valid survey instruments is comprehensively discussed in Saris and Gallhofer (2007). Van Deth (1998) and Hoffmeyer-Zlotnik and Wolf (2003) focus on the omnipresent challenge for cross-national surveys of finding functional equivalent measures in case the concept exists, but the same measurement instrument cannot be used across groups.

Measurement Error

At the measurement error stage, the TSE, as structured by Pennell et al., captures the numerous potential sources of error that may occur during the complete response process. For a respondent, it is not possible to give a valid answer when the question was not properly understood. This may be due to a translation error, making translation one of the crucial sources of measurement error in cross-national, multilingual surveys, where basic questionnaires have to be translated into the national languages. A comprehensive overview of the core features of good practice translation can be found at Behr and Shishido (2016, 271 (table 19.1)) as well as in Mohler et al. (2016). In this context Harkness, Villar and Edwards (2010) discuss the characteristics of a successful translation and its challenges. Behr and Shishido (2016) explore the role of translation for the development of equivalent questions in cross-national studies; and Schneider et al. (2016) focus on the fact that sometimes "translation is not enough", because even seemingly straightforward concepts are socially constructed and consequently linked to institutional structures and context-dependent interpretations (Schneider et al. 2016, 287).

Response distributions may differ due to consciously or unconsciously applied response styles. Much research has been done on response styles (see for an overview Roberts 2016) as well as culture-specific response styles (Marin et al. 1992; Baumgartner and Steenkamp 2001; Yang et al. 2010) and the impact these have on data quality. When conducting cross-national research it should be kept in mind that within some cultures respondents tend to answer more extremely and generally make use of all categories available along a scale, while in other countries respondents tend to answer more moderately (for an overview of studies see Yang 2010, 208f.) without actually holding significantly different attitudes.

Also, question format as well as the specific wording may significantly affect response distributions (Billiet and Matsuo 2012). For example, open questions yield different responses than closed questions and offering a "don't know" category and/or a middle category of a scale lead to different response distributions than if respondents are forced to pick a substantial category.

Previous research has also shown that even the interviewer, just by her/his sheer presence, representing a specific gender, age or ethnic group may influence response behavior (see for example West et al. 2013). Most recently, these effects have been studied by Beullens and Loosveldt (2016) whereby using six rounds of ESS data, they demonstrated that interviewer effects tend to be severely underestimated in research practice.

Survey questions, especially on sensitive topics and when asked personally by an interviewer in a face to face interview, can also lead to *measurement error* caused by the phenomenon of social desirability. Social desirability means respondents answering according to the perceived norm rather than to their true opinions and attitudes. The theory behind this phenomenon has been treated in depth by Tourangeau et al. (2000), Johnson and Van de Vijver (2003) and Lensvelt-Mulders (2008).

Using a self-completion mode for the survey might reduce *measurement error* due to social desirability. However, without the guidance and control of the interviewer there are other types of error which may be produced by the respondent. Any decision for the potential trade-off should be made by country experts, one might say. This, however, would most certainly lead to different modes of interviewing across countries, which might in turn also affect response behavior. If respondents answer differently to questions solely because of the mode of interviewing, we speak of *measurement error* due to mode effects. These effects have been analyzed cross-nationally by Skjåk and Harkness (2003) as well as on the basis of a national mixed-mode survey by Lugtig et al. (2011).

Processing error

Pennell's scheme also includes *processing error* as a potential contributor to measurement error. Processing errors may occur during the data collection process by the interviewers. They may also occur after data has been collected during the data handling processes, such as coding of responses on open questions into certain standard schemes (such as ISCO for occupation or ISCED of education), recoding missing values, output harmonization of variables or matching of cases from different samples. The integration process of the national datasets into the full internationally comparable datasets at the responsible institutions of course also

bears some risk of processing error. At these stages of the development process, systematic quality checks, such as tests for duplicate records (Kuriakose and Robbins 2015; Slomczynski et al. 2017) are highly recommend for detecting and correcting processing error as fast as possible. Expending sufficient resources on staff training and monitoring is the best way to minimize processing error.

Evaluation of cross-national comparability

This is a good point at which to depart from the scope of the TSE and touch on, in my opinion, what is consequentially the next step after categorizing potential survey error: the evaluation of which level of comparability could actually be achieved across country data. For this evaluation, different, well-established test procedures can be applied at different stages of the survey process. A comprehensive overview of the different procedures can be found in Braun and Johnson (2010).

Proceeding in chronological order, the first tests should be pretests of measurement instruments conducted in the design phase of questionnaire development. Ideally, qualitative pretests should be conducted to discover difficulties respondents might have in answering the questions, while quantitative pretests enable testing the data for invariance by means of statistical procedures (Behr et al. 2016). Pretests should be scheduled some time before the main survey and include as many different countries and cultural areas as the budget allows. If modifications are made to the questions, however, they should possibly be pre-tested again (for pretest techniques and their evaluation see Caspar et al. 2016).

Despite all efforts to prevent survey errors, there is no guarantee of data comparability in the final international data releases. However, every researcher has the possibility to empirically test for comparability with different quantitative methods. Multilevel Analysis *ML* (Raudenbush and Bryk 1986) which has become increasingly popular in this respect over the last few years, treats each country as a sample from a larger population. It therefore does not estimate single parameter values for each country to be compared, but assumes a (normal) distribution of parameter values across countries and calculates its variance (Hox et al. 2010). Yet, the most commonly used approaches are probably still Multigroup Confirmatory Factor Analysis *MGCFA* (Jöreskog 1971; Steenkamp and Baumgartner 1998), Item Response Theory *IRT* (Raju et al. 2002), and Latent Class Analysis *LCA* (Kankaraš et al. 2011). The point of departure for all three approaches is a latent variable construct. Testing for data comparability with a latent variable approach is based on the principle of comparing data structures by analyzing the relations between the latent variable and the indicators across groups. The more similar these data structures are, the better the comparability. Davidov et al. (2014) distinguish

these three approaches in detail and explain which technique is the most suitable for what kind of data, depending on the measurement levels of observed and latent variables.

Methodological research is constantly evolving. The above-mentioned approaches are characterized by being rather strict in attesting non-invariance. When large numbers of countries, comprised of many different backgrounds, are involved in models, however, comparability is sometimes questioned a bit too hastily. Taking into account that problem, Asparouhov and Muthén (2013) most recently developed the so-called *alignment optimization, a method* based on the assumption that the amount of non-invariance can be kept to a minimum level. The procedure estimates the factor means while discovering the most invariant model possible (Muthén and Asparouhov 2013). This new method is appealing but still needs more empirical validation (Lomazzi 2017). I therefore decided to adopt a more consolidated approach by conducting the comparability tests via LCA.

The previous sections of this frame-paper for my dissertation project provided an overview on the TSE framework for comparative surveys, as well as the research behind it: on the nature of survey errors, error prevention and methods to assess levels of comparability. The three articles of my dissertation project may be placed in this framework.

1.4 Extended Abstracts of the Three Articles

Title article I: Observing Unexpected Patterns in Cross-National Research: *Blame Data, Theory, or Both? Attitudes toward Redistributive Taxation in Thirty-Three Countries*

The first article, written together with Jonas Edlund[5] deals with the construct of welfarism, more specifically, it examines the relationships between socioeconomic status and attitudes toward redistributive taxation across thirty-three countries using the complete 2006 ISSP *Role of Government IV* data set. We apply a simple rational-choice-inspired homo-economicus hypothesis (H1) proposing that those better off in the socioeconomic hierarchy should have less reason to support state-organized economic redistribution compared to those situated at lower levels in the socioeconomic hierarchy.

H1: The lower the socioeconomic status, the stronger the support for redistributive taxation.

The empirical results demonstrate substantial cross-country variation regarding the correspondence between empirical observations and theoretical expectations. When faced

[5] Umeå University, Department of Sociology

with such tremendous cross-national variation in response patterns, a common strategy among researchers is to question the quality of the data collection procedures for those countries deviating strongly from theoretical expectations. We argue that an observed lack of fit between theory and empirical observations may be rooted in comparability problems related to theory rather than the quality of data collection procedures. Building on previous research, two "cultural distance" hypotheses are formulated, both of which argue that the correspondence between the homo-economicus theory and empirics should indeed vary systematically across countries. The first focuses on the role of the welfare state:

> *H2: The more encompassing the welfare state, the stronger the correlation between socioeconomic status and redistributive taxation in the direction as formulated in H1.*

The second "cultural distance" hypothesis takes into account the assumed dominance of more affluent Western countries in research (Wallerstein 1997).

> *H3: The higher the level of economic development in a country, the stronger the correlation between socioeconomic status and redistributive taxation in the direction as formulated in H1.*

To test these hypotheses a series of predictability measures based on the expected patterns according to the homo-economicus theory were developed. The higher a country's predictability coefficient, the better the theory's explanatory potential.

H1 and *H2* receive considerable empirical support. The relationship between socioeconomic status and support for redistributive taxation is substantially stronger in the wealthy Western welfare states, particularly among those of Northern Europe, than in the less affluent non-Western countries lacking any institutional design reminiscent of a welfare state. The analysis of specific country response patterns, showing strikingly different preferences toward progressive taxation than in other countries, provided no reason to believe that they were rooted either in coverage or translation error.

Contribution to knowledge*:* The study shows that theoretically deviant empirical observations in cross-national research may be related to theory generalization limitations. Specifically, the weak correspondence between the homo-economicus theory and the empirical results observed in most of the non-Western countries indicate that here Western theoretical assumptions are simply not valid. For cross-national research that means caution is warranted when comparing attitudes towards progressive taxation across countries with significantly different welfare states.

Next research steps: The consequent next steps of research are (1) to investigate the value of the homo-economicus theory for more Western as well as non-Western countries for the predictability of attitudes towards redistribution. And (2) test the generalizability of other theories across countries using the predictability measure procedure developed in this study.

Articles II and III leave the research field of welfarism and focus on comparability issues with respect to the theoretical construct of religion:

Title article II: Comparing Religiosity Cross-Nationally - *About Invariance and the Role of Denomination*

The point of departure for the second article is the shared understanding in the international sociological research field of religion that religiosity is a multidimensional and multifaceted phenomenon (see for example Glock 1962; Storm 2009). Factors such as belief, emotional connection to God, spirituality, religious behavior, or institutional affiliation account for different dimensions of religiosity. Individuals may be very religious with respect to one dimension and less so towards another. One dominant research opinion states that religiosity can only satisfactorily be captured by typological measurement approaches since individual combinations of religious factors lead to individual typologies, also called profiles of religiosity (see for example Storm 2009). In previous research, various religious profiles, showing different combinations of religious multidimensionality, have been identified within and across countries. “Believing without belonging” (Davie 1990) might be the most famous profile that was found. However, comparative research has also shown that these observable dominant religious profiles are not necessarily the same across countries (see for example Davie 1994; 2000; Storm 2009 or Voas 2009).

Recent methodological approaches stress the necessity of testing data for measurement invariance before comparison. This is particularly important for cross-national data since respondents are socialized in different cultural contexts, live in different economic situations and speak different languages - all factors that might lead to dissimilar and varied understandings of survey questions and related concepts. To make meaningful comparisons, data must be truly comparable, despite the enormous challenges. (See Vandenberg and Lance 2000; Meuleman and Billiet 2012; Davidov et al. 2014; to mention but a few).

By applying Latent Class Analyses (LCA) including invariance tests on ISSP *Religion III* data from 2008, the article identifies dominant cross-national profiles of religiosity across thirty-five

countries and answers the central research question:

> *Is it possible to conduct a typology including thirty-five countries across the world that produces results which pass an invariance test?*

The nature of a survey conducted across six continents is such that some of the thirty-five countries included in the analysis will inevitably be very different. Differences occur regarding the economic and political situation, recent history, or the Church-State relationship, to list only a few factors. Finding dominant profiles, valid across a large number of different countries therefore appears challenging at best. It is assumed that the level of heterogeneity decreases if the whole set of countries is broken down into smaller groups. Consequently, the second part of the central research question is:

> *If it is not possible to find a typology that proves worldwide validity, how much closer towards invariance can we get by building smaller subgroups?*

For the formation of religious profiles, the religion or denomination individuals belong to is regarded as a crucial factor, since religions and denominations respectively, are based on theologies emphasizing different transcendent or ritual aspects of religiosity (Halman and Draulans 2004). Following a theoretical exploration of what kind of impact belonging to certain religions/denominations might play in individual religious profile formation, the article goes on to examine the impact the denomination has on profile formation in general and tests hypotheses on the specific impact of different denominations in particular. In that vein, the article extends the comparability issue by comparing data not only across nations but also across denominational groups.

The results reveal there is a valid typology worldwide, but that instead of showing specific characteristics, profiles can be ordered on a single latent continuum from low to high according to the general strength of respondents' religiosity. The second finding is that it is almost impossible to establish full homogeneity (invariance) across countries with a comprehensive set of variables measuring religiosity. Partial homogeneity, however, can be achieved. Furthermore, against all theoretical expectations, the analyses imply that denomination is not a crucial impact factor for religious profile formation.

Contribution to knowledge: For international survey research the finding that religiosity can actually be validly measured and compared across countries with the data available from international comparative surveys is valuable knowledge. It was not realistic to expect similar intercepts across the data that indicate equal general levels of religiosity in societies across the world. But similar slopes, assuring comparability of the data could be established and the data, therefore, can be used for comparative analyses. The results though, do question the

usefulness of typological measurement approaches for measuring religiosity by showing that nominal latent classes resulting from analyses can be ordered on one continuum for low to high levels of religiosity. These results raise the idea that continuous measures (which are usually less complicated to apply) are at least equally well suited.

Next research steps: Based on these findings future research should make the effort to disentangle possible interaction effects of country-specific macro indicators and the dominant denomination in the countries to retest for a possible influence on individual religious profile development.

Title article III: Is there a Continuous Way to God? *A Cross-National Evaluation of the Indicator "Self-Assessed Level of Religiosity" as a Continuous Single-Item Measure of Religiosity*

The third article is also based on the understanding of religiosity as a multifaceted, multidimensional phenomenon. International survey data nevertheless usually contain only one or two classical indicators for the measurement of religiosity. For lack of a greater variety of indicators, multivariate analyses have little choice other than to include only one of these usually continuous indicators, or a composite index of the few indicators available to explain the research subject. Given the complex nature of religiosity, this inevitably raises the question:

> *How well can a single continuous measure explain a multifaceted, multidimensional phenomenon, such as religiosity?*

Previous research has examined the validity of *church attendance* combined with *membership in a denomination* as a measure for church religiosity in a Christian context (Jagodzinski and Dobbelaere 1995). However, church religiosity is not the same as individual religiosity which may be dissociated from classical churches – a type of religiosity actually on the rise in many countries (see, for instance, Beckford 2003). By contrast, a personal indicator measuring the individually assessed, cross-dimensional intensity of religiosity seems more promising for future research.

Utilizing the 2008 ISSP *Religion III* data, this article examines the validity and predictive potential of the single continuous indicator "self-assessed level of religiosity" (SAR) as a proxy for the measurement of religiosity in international comparative research. The indicator is compared with two other measurement instruments for the analyses. These are constructed on a multi-item basis: one continuous, as the SAR measure; and one categorical. The categorical measurement instrument contains latent classes of religiosity produced by Latent

Class Analyses (LCA). It is characterized by the conditional probabilities of seven indicators addressing different dimensions of religiosity. The same indicators are used to construct the second multivariate, but continuous measure: a principle component resulting from Structural Equation Modelling (SEM).

All measures are tested for thirty-five countries. Test procedure steps include (1) evaluating the continuous SAR indicator's predictive potential, (2) comparing the results with the likewise continuous, but multi-item SEM - and the multi-item, but categorical LCA approaches, and (3) analyzing whether the predictive potential of the three measures varies across thirty-five countries.

The results reveal a rather unexpected, generally convincing performance of the proxy. However, they also make quite clear that the measures' performance is context-dependent to a large extent. The outcomes further confirm what the results already suggested in article II by detecting a structure of nominal latent classes that can be ordered on a continuum for low to high levels of religiosity: continuous measures of religiosity generally do not have a weaker predictive potential for the measurement of religiosity than categorical measures.

Contribution to knowledge: These results have direct implications for survey questionnaire development and practical analysis. The proof that SAR is an acceptable proxy for measurement of religiosity in the vast majority of countries recommends it for implementation in surveys. This article designates those countries, especially ones with generally high levels of public religiosity, in which SAR results should be treated carefully, nevertheless. To capture slight variances in these cases it is advisable to use a multi-item measure. If appropriate items are available, SEM is more reasonable to use than LCA. Exceptions to that "rule" appear to be due to the dependent subject matter rather than country characteristics.

Next research steps: Against the backdrop of this research, it will be interesting to test the different measures of religiosity beyond the scope of Christianity. A starting point might be to take Asian countries and religions where Hinduism and Buddhism are prevalent, and which record a considerable number of members across ISSP surveys. A second step, which is quite attractive, involves using cognitive interviews in an attempt to get at the basis of what dimension(s) of religiosity people (across countries and religions) actually address when asked how religious they are. It might be prudent to slightly delay this endeavor for the moment as the data for the upcoming ISSP survey *Religion IV* will be available as of 2020.

1.5 Conclusion

My dissertation project underlines the conclusions of Sir Jowell: "[…] measurement difficulties faced by cross-national studies are hardly a reason for abandoning them" (Jowell 1998, 173). With some circumspection in the handling of comparative data, analyses produce valid and comparable results.

In my dissertation project, I look at selected international data comparability issues which can be assigned to different elements of the *Total Survey Error* (TSE) paradigm. As I have shown in the literature review of this frame-paper, other researchers treated other TSE elements theoretically and empirically. The big challenge is to be sensitive to these issues potential pitfalls and keep the potential sources of error in mind when analyzing the data. This sensitivity is the basis for correctly evaluating the most potential error sources that should be examined in depth before conclusions are drawn. Reading the data documentation carefully is a first, indispensable step in this evaluation process. Nevertheless, empirically testing comparability of the data selected for use remains inevitable.

If empirical observations and theoretical expectations differ substantially, the key question then becomes: blame the data or blame the theory? An unequivocal answer to that question is the exception rather than the rule. What is certain, however, is that both need to be scrutinized. Often enough, insights gained by examining assumed purely methodological aspects may help our understanding of theory. What is discovered as a translation mistake, for example, may in fact reflect a general, culture-based misunderstanding of the concept. Therefore, it might not be wise to separate methodological and theory research too strictly from each other and ensure that the lessons learned from this research are in turn used to improve social theory.

1.6 References

Asparouhov, Tihomir, and Bengt Muthén. 2013. "Multiple-Group Factor Analysis Alignment." *Structural Equation Modeling: A Multidisciplinary Journal* 21 (4): 495-508.

Aspinall, Peter J. 2007. "Approaches to Developing an Improved Cross-National Understanding of Concepts and Terms Relating to Ethnicity and Race." *International Sociology* 22 (1): 41-70.

Barnes, Samuel H., and Max Kaase. 1979. "Introduction". In *Political Action: Mass Participation in Five Western Democracies*, edited by Samuel H. Barnes, and Max Kaase, 13-26. Beverly Hills, London: SAGE.

Baumgartner, Hans, and Jan-Benedict E.M. Steenkamp. 2001. "Response Styles in Marketing Research: A Cross-National Investigation." *Journal of Marketing Research* 38 (2): 143-156.

Beckford, James A. 2003. *Social Theory and Religion*. Cambridge: Cambridge University Press.

Behr, Dorothée, Michael Braun, and Brita Dorer. 2016. "Measurement Instruments in Cross-National Surveys." *GESIS Survey Guidelines*. Mannheim: GESIS - Leibniz Institute for the Social Sciences. Accessed September 17, 2017. http://dx.doi.org/10.15465/gesis-sg_en_006.

Behr, Dorothée, and Kuniaki Shishido. 2016. "The Translation of Measurement Instruments for Cross-Cultural Surveys." In *The SAGE Handbook of Survey Methodology*, edited by Christof Wolf, Dominique Joye, Tom W. Smith, and Yang-chih Fu, 269-287. Los Angeles: SAGE.

Beullens, Koen, and Geert Loosveldt. 2016. "Interviewer Effects in the European Social Survey." *Survey Research Methods* 10 (2): 103-118.

Biemer, Paul P. 2010. "Total Survey Error. Design, Implementation, and Evaluation." *Public Opinion Quarterly* 74 (5): 817–848.

Biemer, Paul P., Edith de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker, and Brady T. West. 2017. *Total Survey Error in Practice*. Hoboken: Wiley.

Billiet, Jaak, and Hideko Matsuo. 2012. "Non-Response and Measurement Error." In *Handbook of Survey Methodology for Social Sciences*, edited by Lior Gideon, 149-178. New York: Springer.

Billiet, Jaak. 2016. "What does Measurement Mean in a Survey Context?", In *The SAGE Handbook of Survey Methodology*, edited by Christof Wolf, Dominique Joye, Tom W. Smith, and Yang-chih Fu, 193-210. Los Angeles: SAGE.

Blasius, Jörg, and Victor Thiessen. 2012. *Assessing the Quality of Survey Data*. London: Sage.

Braun, Michael, Dorothée Behr, and Lars KaczmirekL. 2012. "Assessing Cross-National Equivalence of Measures of Xenophobia: Evidence from Probing in Web Surveys." *International Journal of Public Opinion Research* 25 (3): 383–395.

Braun, Michael, and Timothy P. Johnson. 2010. “An Illustrative Review of Techniques for Detecting Inequivalences.” In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith, 375-393. Hoboken: Wiley.

Caspar, Rachel, Emilia Peytcheva, Ting Yan, Sunghee Lee, Mingnan Liu, and Menyao Hu. 2016. “Pretesting.” *Cross-Cultural Survey Guidelines*. Accessed November 16, 2017. http://ccsg.isr.umich.edu/index.php/chapters/pretesting.

Davidov, Eldad, Bart Meuleman, Jan Cieciuch, Peter Schmidt, and Jaak Billiet. 2014. “Measurement Equivalence in Cross-National Research.” *Annual Review of Sociology* 40: 55–75.

Davie, Grace. 2000. *Religion in Modern Europe: A Memory Mutates*. Oxford: Oxford University Press.

Davie, Grace. 1994. *Religion in Britain since 1945: Believing without Belonging.* Oxford: Blackwell.

Davie, Grace. 1990. “Believing without Belonging - Is This the Future of Religion in Britain?” *Social Compass* 37 (4): 455-469.

Drake, H. Max. 1973. “Research Method or Culture-Bound Technique? Pitfalls of Surveay Research in Africa.” In *Survey Research in Africa: Its Applications and Limits*, edited by William M. O’Barr David H. Spain, and Mark A. Tessler, 58-69. Evanston: Northwestern University Press.

GESIS. 2017. “Overview of Comparative Surveys Worldwide.” Accessed December 10, 2017. https://www.gesis.org/en/services/data-analysis/more-data-to-analyze/overviews/overview-of-comparative-surveys-worldwide/.

Glock, Charles Y. 1962. “On the Study of Religious Commitment.” *Religious Education* 57: 98-110.

Groves, Robert M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.

Groves, Robert M., Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. Hoboken: Wiley.

Häder, Sabine, and Siegfried Gabler. 2003. “Sampling and Estimation.” In *Cross-Cultural Survey Methods*, edited by Janet A. Harkness, Fons J. R. van de Vijer, and Peter Ph. Mohler, 117–136. Hoboken: Wiley.

Halman, Loek, and Veerle Draulans. 2004. “Religious Beliefs and Practices in Contemporary Europe.” In *European Values at the Turn of the Millennium*, edited by W. Arts and L. Halman, 283-316. Leiden: Brill.

Harkness, Janet A., Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith. 2010*. Survey Methods in Multinational, Multiregional, and Multicultural Contexts.* Hoboken: Wiley.

Harkness, Janet A., Ana Villar, and Brad Edwards. 2010. “Translation, Adaptation, and Design.” In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith, 117-140. Hoboken: Wiley.

Heath, Anthony F., Silke L. Schneider, and Sarah Butt. 2016. “Developing a Measure of Socio-Cultural Origins for the European Social Survey.” *GESIS Papers 2016/16*. Cologne: GESIS.

Hoffmeyer-Zlotnik, Jürgen H.P., and Uwe Warner. 2011. “Measuring Occupation and Labour Status in Cross-National Comparative Surveys.” *GESIS Schriftenreihe* 7. Bonn: GESIS.

Hoffmeyer-Zlotnik, Jürgen H.P., and Uwe Warner. 2014. *Harmonising Demographic and Socio-Economic Variables for Cross-National Comparative Survey Research*. Dordrecht: Springer.

Hoffmeyer-Zlotnik, Jürgen H.P., and Christof Wolf. 2003. “Comparing Demographic and Socioeconomic Variables across Nations: Synthesis and Recommendations.” In *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*, edited by Jürgen H.P. Hoffmeyer-Zlotnik and Christof Wolf, 389-406. New York: Kluwer.

Hox, Joop J. 1997. “From Theoretical Concepts to Survey Question.” In *Survey Measurement and Process Quality*, edited by Lars E. Lyberg, Paul P. Biemer, Martin Collins, Edith de Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin, 47-69. New York: Wiley.

Hox, Joop J., Edith de Leeuw, and Mathieu J.S. Brinkhuis. 2010. “Analysis Models for Comparative Surveys.” In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith, 395-418. Hoboken: Wiley.

Jagodzinski, Wolfgang, and Karel Dobbelaere. 1995. “Secularization and Church Religiosity.” In *The Impact of Values*, edited by Jan W. van Deth, and Elinor Scarbrough, 76-119. Oxford: Oxford University Press.

Johnson, Timothy P., and Michael Braun. 2016. “Challenges of Comparative Survey Research.” In *The SAGE Handbook of Survey Methodology*, edited by Christof Wolf, Dominique Joye, Tom W. Smith, and Yang-chih Fu, 41-53. Los Angeles: SAGE.

Johnson, Timothy P., and Fons van de Vijver. 2003. “Social Desirability in Cross-Cultural Research.” In *Cross-Cultural Survey Methods*, edited by Janet A. Harkness, Fons van de Vijver, and Peter Ph. Mohler, 195-204. Hoboken: Wiley.

Jöreskog, Karl G. 1971. “Simultaneous Factor Analysis in Several Populations.” *Psychometrika* 36 (4): 409–426.

Jowell, Roger. 1998. “How Comparative Is Comparative Research?” *American Behavioral Scientist* 42 (2): 168-177.

Kankaraš, Miloš, Guy B.D. Moors, and Jeroen K. Vermunt. 2011. “Testing for Measurement Invariance with Latent Class Analysis.” In *Cross-Cultural Analysis*, edited by Eldad Davidov, Peter Schmidt, and Jaak Billiet, 359–384. New York: Routledge.

Kerlinger, Fred N. 1973. *Foundations of Behavioral Research.* London: Holt, Rinehart and Winston.

Kuriakose, Noble, and Michael Robbins. 2015. “Falsification in Surveys: Detecting Near Duplicate Observations”. *Arab Barometer Working Paper* 2. Accessed April 7, 2016. http://www.arabbarometer.org/content/wp2-noble-kuriakose-and-michael-robbins-detecting-near-duplicates-survey-research.

Laerd Dissertation. 2016. “Constructs in Quantitative Research.” Accessed December 13, 2017. http://dissertation.laerd.com/constructs-in-quantitative-research.php.

Lensvelt-Mulders, Gerty J.L.M.. 2008. “Surveying Sensitive Topics.” In *International Handbook of Survey Methodology*, edited by Edith de Leeuw, Joop J. Hox, and Don A. Dillman, 461-478. New York: Erlbaum.

Lomazzi, Vera. 2018. *Forthcoming* “Using Alignment Optimization to Test the Measurement Invariance of Gender Role Attitudes in 59 Countries*.” methods, data, analyses.*

Lugtig, Peter J., Gerty J.L.M. Lensvelt-Mulders, Remco Frerichs, and Assyn Greven. 2011. “Estimating Nonresponse Bias and Mode Effects in a Mixed-Mode Survey.” *International Journal of Market Research* 53 (5): 669–686.

Marin, Gerardo, Raymond J. Gamba, and Barbara V. Martin. 1992. “Extreme Response Style and Acquiescence among Hispanics.” *Journal of Cross-Cultural Psychology* 23 (4): 498-509.

Matsuo, Hideko, Jaak Billiet, Geert Loosveldt, Frode Berglund, and Øyven Kleven. 2010. “Measurement and Adjustment of Non-Response Bias Based on Non-Response Surveys: The Case of Belgium and Norway in the European Social Survey Round 3.” *Survey Research Methods* 4 (3): 165-178.

Meuleman, Bart, and Jaak Billiet. 2012. “Measuring Attitudes toward Immigration in Europe: The Cross-Cultural Validity of the ESS Immigration Scales.” *Ask: Research and Methods* 21 (1): 5–29.

Mohler, Peter, Brita Dorer, Julie de Jong, and Mengyao Hu. 2016. “Translation: Overview.” *Guidelines for Best Practice in Cross-Cultural Surveys*. Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan. Accessed September 26, 2017.http://www.ccsg.isr.umich.edu/index.php/chapters/translationchapter/translation-overview/.

Muthén, Bengt, and Tihomir Asparouhov. 2013. “New Methods for the Study of Measurement Invariance with Many Groups.” Accessed December 14, 2017. https://www.statmodel.com/download/PolAn.pdf.

Ortmanns, Verena, and Silke L. Schneider. 2016. "Harmonization Still Failing? Inconsistency of Education Variables in Cross-National Public Opinion Surveys." *International Journal of Public Opinion Research* 28 (4): 562-582.

Pennell, Beth-Ellen, Kristen Cibelli Hibben, Lars E. Lyberg, Peter Ph. Mohler, and Gelaye Worku. 2017. “A Total Survey Error Perspective on Surveys in Multinational, Multiregional, and Multicultural Contexts.” In *Total Survey Error in Practice*, edited by Paul. B. Biemer, Edith de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker, and Brady T. West, 179-202. Hoboken: Wiley.

Raju, Nambury S., Larry J. Laffitte, and Barbara M. Byrne. 2002. “Measurement Equivalence: A Comparison of Methods Based on Confirmatory Factor Analysis and Item Response Theory.” *Journal of Applied Psychology* 87 (3): 517–529.

Raudenbush, Stephen, and Anthony S. Bryk. 1986. “A Hierarchical Model for Studying School Effects.” *Sociology of Education* 59 (1): 1-17.

Roberts, Caroline. 2016. “Response Styles in Surveys: Understanding their Causes and Mitigating their Impact on Data Quality.” In *The SAGE Handbook of Survey Methodology*, edited by Christof Wolf, Dominique Joye, Tom W. Smith, and Yang-chih Fu, 579-596. London: SAGE.

Saris, Willem E., and Irmtraud N. Gallhofer. 2007. *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Hoboken: Wiley.

Schneider, Silke L. 2009. *Confusing Credentials: The Cross-Nationally Comparable Measurement of Educational Attainment.* Oxford: University of Oxford.

Schneider, Silke L., Dominique Joye, and Christof Wolf. 2016. "When Translation Is Not Enough: Background Variables in Comparative Surveys." In *The SAGE Handbook of Survey Methodology*, edited by Christof Wolf, Dominique Joye, Tom W. Smith, and Yang-chih Fu, 288-307. London: SAGE.

Skjåk, Knut Kalgraff, and Janet A. Harkness. 2003. “Data Collection Methods.” In *Cross-Cultural Survey Methods*, edited by Janet A. Harkness, Fons J.R. van de Vijver, and Peter Ph. Mohler, 179-194. Hoboken: Wiley.

Slomczynski, Kazimierz M., Przemek Powalko, and Tadeusz Krauze. 2017. “Non-Unique Records in International Survey Projects: The Need for Extending Data Quality Control.” *Survey Research Methods* 11 (1): 1-16.

Smith, Tom W. 2011. “Refining the Total Survey Error Perspective.” *International Journal of Public Opinion Research* 23 (4): 464-484.

Smith, Tom W., and Yang-chih Fu. 2016. “The Globalization of Surveys.” In *The SAGE Handbook of Survey Methodology*, edited by Christof Wolf, Dominique Joye, Tom W. Smith, and Yang-chih Fu, 680-692. London: SAGE.

Smith, Tom W. 2017. “Using the Total Survey Error Paradigm to Improve Cross-National Research on Religion” in *Faithful Measures: New Methods in the Measurement of Religion*, edited by Roger Finke and Christopher D. Bader. New York: New York University Press.

Steenkamp, Jan-Benedict E.M., and Hans Baumgartner. 1998. “Assessing Measurement Invariance in Cross-National Consumer Research.” *Journal of Consumer Research* 25 (1): 78–90.

Storm, Ingrid. 2009. “Halfway to Heaven: Four Types of Fuzzy Fidelity in Europe.” *Journal for the Scientific Study of Religion* 48 (4): 702-718.

Survey Research Center. 2016. *Guidelines for Best Practice in Cross-Cultural Surveys: Full Guidelines*. Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan. Accessed September 26, 2017. http://www.ccsg.isr.umich.edu/.

Tourangeau, Roger, Lance J. Rips, and Kenneth A Rasinsk. 2000. *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Van Deth, Jan W. 1998. “Equivalence in Comparative Political Research.” In *Comparative Politics. The Problem of Equivalence,* Routledge: London/New York 1-19.

Vandenberg, Rober J., and Charles E. Lance. 2000. “A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research.” *Organizational Research Methods* 3 (1): 4-70.

Voas, David. 2009. “The Rise and Fall of Fuzzy Fidelity in Europe.” *European Sociological Review* 25 (2): 155–168.

Wallerstein, Immanuel. 1997. “Eurocentrism and Its Avatars: The Dilemmas of Social Science.” *Sociological Bulletin* 46 (1): 21-39.

Weisberg, Herbert F. 2005. *The Total Survey Error Approach: A Guide to the New Science of Survey Research.* Chicago: University of Chicago Press.

West, Brady T., Frauke Kreuter, and Ursula Jaenichen. 2013. “’Interviewer’ Effects in Face-to-Face Surveys: A Function of Measurement Error or Nonresponse?” *Journal of Official Statistics* 29 (2): 277-297.

Wolf, Christof, Dominique Joye, Tom W. Smith, and Yang-chih Fu. 2016. *The SAGE Handbook of Survey Methodology.* London: SAGE.

Yang, Yongwei, Janet A. Harkness, Tzu-Yun Chin, and Ana Villar. 2010. “Response Styles and Culture.” In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith, 203-226. Hoboken: Wiley.

This is an Accepted Manuscript of an article published by Taylor & Francis in *International Journal of Sociology* on 07 Dec 2015, available online: http://www.tandfonline.com/doi/full/10.1080/00207659.2015.1098272.

2. Observing Unexpected Patterns in Cross-National Research: Blame Data, Theory, or Both? *Attitudes toward Redistributive Taxation in Thirty-Three Countries*[6]

Abstract

This article examines the relationships between socioeconomic status and attitudes toward redistributive taxation across 33 countries using the complete International Social Survey Programme (ISSP) 2006 data set. We apply a simple rational-choice-inspired homo-economicus hypothesis proposing that those better off in the socioeconomic hierarchy should have less reason to support state-organized economic redistribution compared to those situated at lower levels in the socioeconomic hierarchy. The empirical results demonstrate substantial cross-country variation regarding the correspondence between empirical observations and theoretical expectations. When faced with such tremendous cross-national variation in response patterns, a common strategy among researchers is to question the quality of the data collection procedures for those countries deviating strongly from theoretical expectations. The strategy chosen in this study is, however, different. The main argument is that an observed lack of fit between theory and empirical observations may be rooted in problems related to theory rather than the quality of data collection procedures. Building on previous research, two "cultural distance" hypotheses are formulated, both of which argue that the correspondence between the homo-economicus theory and empirics should indeed vary systematically across countries. The first focuses on the role of the welfare state and the second on the level of economic affluence and associated scientific dominance. Both hypotheses receive considerable empirical support. The relationship between socioeconomic status and support for redistributive taxation is substantially stronger in the wealthy Western welfare states—particularly among those of Northern Europe—than in the poor non-Western countries lacking any institutional design reminiscent of a welfare state.

Keywords: attitudes; international comparative research; ISSP; latent class analysis; taxation

[6] Insa Bechert, and Jonas Edlund. 2015. *International Journal of Sociology*, 45(4): 327-347.

2.1 Introduction

International comparisons are seldom easy; even seemingly straightforward concepts such as income distribution or gross domestic product (GDP) can be very difficult to compare across countries. The difficulties encountered are no less significant when it comes to attitudes and values. This study deals with attitudes toward taxation across 33 countries. A very simple—some may even say "common sense"—rational choice-based theory on the relationships between socioeconomic position and preferences for state-organized economic redistribution says that the higher the socioeconomic position, the lower the probability for supporting redistributive taxation.

Now, let us assume that this hypothesis is empirically tested using data from a fairly large number of countries, and that the results indicate tremendous cross-national variation. In some countries the expected pattern is clearly visible, while in other countries the observed relationship is close to zero, or even worse, inverted. In other words, the predictive power of the theory differs substantially across countries. Faced with such a result, the dilemma to be solved is the following: is the lack of fit between theory and the empirical results that can be observed in some countries rooted in data of insufficient quality or to qualities related to the theory itself, that is, that the theory is more valid in some countries than in others?

While researchers may respond differently to this dilemma, the first step is often to focus on the data. Some may take the easiest way out, simply not using the countries where the most unexpected results appear. In fact, based on informal conversations with scholars involved in comparative attitudinal research, this response is far from uncommon. Another response is to report the statistical properties of the observed associations and basically argue that the data for the deviant countries are of bad quality (Blasius and Thiessen 2006). A third response is to examine the quality of data in-depth by applying different measures, such as translation checks and sample coverages.[7] Another strategy is to redefine and modify theory. In this study, we will approach the dilemma mainly from this "theory quality" perspective.[8]

Our basic argument is that we should indeed expect that the strength of the relationship between attitudes toward redistributive taxation and socioeconomic position varies systematically across countries. In short, the relationship should be more pronounced in large welfare states compared to small welfare states. In countries lacking any institutional design

[7] An example of an empirically generated theory based on a translation mistake was recently reported by Charles Kurzman in the Washington Post (Kurzman 2014; for the translation check procedure, see Mellon 2011).

[8] Having said that, we will of course not completely ignore focusing on measurement errors related to data. However, we will restrict our attention to the countries displaying the most aberrant response patterns.

reminiscent of a welfare state, the discrepancy between empirics and theoretical expectations should be particularly pronounced. Furthermore, the relationship is expected to be stronger among the wealthy industrialized countries than in the less prosperous countries outside the Western World. These hypotheses are tested on data covering 33 countries collected in 2006 within the framework of the International Social Survey Programme (ISSP).

2.2 Socioeconomic Status and Attitudes toward Redistributive Taxation: Why should Relationships Differ Systematically across Countries?

2.2.1 A Western Bias in Cross-National Research?

Wallerstein (1997) argues that the academic field of the social sciences is a product of the modern world-system of which the choice of subject matter, its theorizing, its methodology, and its epistemology reflect the context in which it was born. It is true that all large contemporary international social surveys have their roots in the Western World—Europe or the United States. Therefore, the leading and well-established theories as well as the corresponding measurements tend to be designed from a Western perspective.[9] Some of the problems associated with applying the same measurements across countries have been recognized by methodologically oriented scholars.[10] Of particular interest in the context of this article is the cultural distance hypothesis formulated by Blasius and Thiessen: "The greater the cultural distance between the origin of a survey instrument and the groups being investigated, the more compromised the data quality and comparability is likely to be" (2012, 5).

We find the concept of cultural distance potentially fruitful. However, there are two issues that need to be discussed. First, while the hypothesis proposes that problems of comparability may be related to data of insufficient quality and/or to nonvalid theoretical assumptions, the research upon which the hypothesis is founded has principally emphasized explanations relating to the quality of data and not on possible errors/fallacies related to theory (Blasius and Thiessen 2006; 2012). Second, "culture" is a notoriously slippery concept. There simply is no general and widely accepted understanding of the concept. However, given the research field within which the cultural distance hypothesis was developed, it seems appropriate to concentrate on theoretically relevant country characteristics that can be seen as "cultural

[9] The attitudinal items used in this study have also been constructed through a Western lens. They actually appeared in the first ISSP survey back in 1985, many years before countries from the non-Western world joined the program.

[10] For a broad overview of methodologically related issues in comparative research, see Harkness et al. (2010a).

markers," defining the degree of similarity/difference between countries.

In our search for such indicators, we have benefitted from insights provided by the historical institutionalism school (Rothstein and Steinmo 2002; Steinmo, Thelen, and Longstreth 1992) as well as from world-systems theory (Wallerstein 1997), and modernization theory (Wilensky 2002). As will be elaborated below, we have selected two types of country-level indicators: the size of the welfare state and the level of economic development.

2.2.2 Explaining Cross-National Variation: The Roles of the Welfare State and Economic Development

Let us begin with the assumed general negative relationship between socioeconomic status and support for state-organized economic redistribution. Following the theory specified by the power resources approach (Korpi 2006; Korpi and Palme 2003), it is suggested that those in weak bargaining positions in the labor market will prefer to locate societal bargaining in parliamentary politics compared to staying within a market relationship. In terms of power, the democratic principle of "one person–one vote" in politics never did and still does not have a counterpart when it comes to relations between labor and capital in Western societies.

Furthermore, if parliamentary representation is obtained, it is expected that those with relatively few market-derived resources will favor redistribution of income by means of state-organized policy, while those in more privileged positions will prefer a larger role for the market–property nexus in distributive processes. From a rational choice point of view, those at the lower end of the socioeconomic status ladder should have stronger reasons for supporting state-organized economic redistribution compared to those located at the upper level. In other words, support for redistribution should come from the poor and resistance from the rich. Corneo and Grüner call this the "homo-economicus effect" (2002, 85), which is also the basic assumption of our first hypothesis (H1).

> *Hypothesis 1: The lower the socioeconomic status, the stronger the support for redistributive taxation.*

Why should we expect that the relationships between socioeconomic position and support for state-organized economic redistribution vary systematically across countries?[11] It is important to recognize the underlying assumptions in the homo-economicus approach: that citizens in general perceive the state-organized systems of taxation and social spending as a class-

[11] This section builds on arguments and empirical applications provided by Edlund (2007), Edlund and Sevä (2013), and Edlund and Lindh (2015).

relevant and efficient apparatus for leveling out market-generated economic inequalities. We propose that the validity of these assumptions differs across countries. In short, the probability that these conditions are fulfilled is significantly higher in a large extensive welfare state context, such as in the encompassing Scandinavian welfare states, compared to in a less extensive welfare state, for example, the American or the Australian welfare states, let alone in countries lacking any institutional setup reminiscent of a welfare state. Why?

The starting point is that the political-institutional framework surrounding citizens in a given context (e.g., country) is of substantial importance for structuring citizens' perceptions, preferences, and behaviors (Mettler and Soss 2004; Soss and Schram 2007; Svallfors 2007). As transmitters of resources and carriers of norms, state policy arrangements can be understood as a filter that delineates both opportunities and constraints, and through which people perceive the world (Öun 2012). In relation to the topic in this study, the following arguments can be applied.

First, the redistributive capacity of the welfare state is determined by the tax system and social spending priorities. The redistributive capacity of the tax system is primarily determined by the size of the total tax revenue. Also, the larger the budget available for social spending, the larger the redistributive capacity of the state (Åberg 1989; Edlund 1999; Edlund and Åberg 2002; OECD 1990).

Empirical evidence clearly shows that the larger the welfare state—irrespective of measures used—the higher the redistributive capacity of the welfare state. Different types of output indicators that have been applied for capturing the size of the welfare state—"total tax revenue/ GDP" or "total social spending/GDP"—are all highly correlated with more direct measures of welfare policy design. The higher the prevalence of universal/encompassing social insurance programs, the higher the levels of taxation and social spending. And, importantly, all of these measures are strongly correlated with the redistributive capacity of the welfare state (Edlund 2007; Edlund and Lindh 2015; Korpi and Palme 1998).

Second, in a large encompassing welfare state, the systems of taxation and social spending involve a comparatively larger proportion of the citizenry—and their resources—than in relatively residual welfare states. Thus, the size of the welfare state determines not only its redistributive capacity but also the extent of its influence on citizens' everyday lives: citizens in more encompassing welfare states pay a larger share of their income in taxes and their livelihood is also more dependent on services and social protection provided by the welfare state (Edlund 2007). For these reasons, citizens in encompassing welfare states are likely to develop a stronger sense of "ownership" of the state and perceive stronger incentives for

being politically involved than citizens in meager welfare states (Persson and Rothstein 2015). To put it somewhat differently, rational voters should be interested in both knowing and trying to influence how tax money is collected and spent—and their interests in policy changes (toward a less ambitious or more ambitious welfare state depending on their location in the class structure) should be stronger among citizens living in a large welfare state compared to those living in a small one.

Third, arguments similar to the above apply to the extent to which the welfare state should attract attention from organized interests, such as political parties. Organized interests are important for underpinning collective memories and worldviews among citizens (Rothstein 2000). By framing public debate and by offering citizens mental tools and guidelines for interpretations of "problems of" and "solutions to" any given issue, organized interests attempt to gain public acceptance of and support for needed action and are, therefore, powerful forces in the processes of attitude formation (Edlund 2007; Svallfors 2006).

In sum, the strength of the link between socioeconomic position and attitudes toward state-organized redistribution is shaped by the redistributive capacity of the welfare state as well as the centrality of the welfare state in political articulation both at the level of organized interests and among ordinary citizens—factors that are all influenced by the size of the welfare state. Based on the arguments above, dealing with institutional feedback effects on public opinion, our second hypothesis (H2) proposes that the predictive power of the above mentioned rational choice-based homo-economicus theory will increase with the size of the welfare state. In countries lacking any institutional design reminiscent of a welfare state, the discrepancy between empirics and theoretical expectations will be particularly pronounced:

> *Hypothesis 2: The more encompassing the welfare state, the stronger the correlation between socioeconomic status and redistributive taxation in the direction as formulated in H1.*

Let us now return to the essential implication of the arguments provided by world-systems theory referred to above, regarding the assumed Western-centered dominance in research (Wallerstein 1997). Using the terminology of Wallerstein, it can be hypothesized that the predictive power of the homo-economicus theory should be stronger among the core countries (the wealthy industrial societies in the Western world) compared to the semiperiphery and periphery countries (the less developed countries located outside the Western world). However, in the perspective of this article, we believe that this three-category classification schema is somewhat too broad and therefore not well-suited as an empirical measure because it is insensitive to systematic variation within the groups of core,

semiperiphery, and periphery countries. We find it more fruitful to use a more fine-grained measure of societal development than is offered by the world-systems theory. In our view, it is more appropriate to use the standard indicator of a country's level of industrialization/economic development (Wilensky 2002): GDP per capita. Therefore, the hypothesis is stated as follows:

Hypothesis 3: The higher the level of economic development in a country, the stronger the correlation between socioeconomic status and redistributive taxation in the direction as formulated in H1.

2.3 Data, Methods, and Variables

The individual level data come from the 2006 Role of Government module of the International Social Survey Programme (ISSP). Each country has provided a national representative random sample of the adult population with a minimum of 1,000 cases.[12] All countries included in the international ISSP Role of Government IV data file are included in the analysis. They vary greatly in relation to geographical belonging, cultural and historical backgrounds, levels of affluence, and welfare state institutional configurations. The Northern countries of Denmark, Finland, Sweden, and Norway are all characterized by high levels of affluence and encompassing large welfare states. Germany, Switzerland, France, and the Netherlands represent affluent continental West European welfare states. The affluent Anglo-American countries: the United States, Canada, New Zealand, Australia, and Great Britain, are, compared to the above countries, characterized by having rather meager welfare states. While all of the above-mentioned countries are well-known research objects within the field of welfare state research (e.g., Esping-Andersen 1992; Korpi and Palme 1998), our analysis also covers a number of countries that are less often used in comparative research. From the Western part of Europe, we include the less affluent South European countries of Spain and Portugal, as well as Ireland—all characterized by rather limited welfare states and strong influences of the Catholic Church (Ferrera 1996). When it comes to the former communist countries of Eastern Europe, data are available for Poland, Slovenia, the Czech Republic, Latvia, Hungary, Russia, and Croatia; from the South American continent: Chile, Uruguay, the Dominican Republic, and Venezuela; from Asia: South Korea, Japan, the Philippines, Taiwan, and Israel. And finally, the only country from the African continent for which data are available is South Africa.

[12] For detailed information on national sampling procedures, data collection, and response rates see the ISSP 2006 Study Monitoring Report (Scholz et al. 2008).

The measures of attitudes toward taxation are based on the following two items: *Generally, how would you describe taxes in [country] today? (We mean all taxes together, including wage deductions, income tax, tax on goods and services, and all the rest.)*

For those with high incomes, are taxes …

For those with low incomes, are taxes …

Answer scale: much too high/too high/about right/too low/much too low?

Both variables have been recoded into three response categories: *Taxes are too high; Taxes are about right; Taxes are too low.*

To make sense of the different combinations of responses that people may give to the two above items, it is appropriate to use a method that has the ability to identify the most common or dominant patterns that exist in the data. For this purpose we have decided to use latent class analysis (LCA). LCA is particularly well-suited for analyzing nonlinear relationships between categorical variables. The assumption of local independence is fundamental to LCA: the total associations between observed variables are entirely explained by an unobserved (latent) variable. LCA distinguishes common preference constellations in the data, and all respondents sharing similar preferences-structures will be allocated to a specific cluster (Hagenaars and Halman 1989). If two dominant types exist among citizens—for example, those preferring a more progressive distribution of taxes vis-à-vis those wanting a more regressive distribution of taxes—a two-cluster model will fit the data. If citizens can be divided into three groups, a three-cluster model will be selected, and so on. The number of dominant preference constellations (clusters) can be determined by applying different model-fit statistics. A very desirable feature of LCA is its ability to calculate the probability for each individual to belong to each cluster, making it possible to estimate the impact of various background variables on individual-level cluster membership probabilities (Magidson and Vermunt 2003).[13] These probabilities will be used as continuous dependent variables in forthcoming ordinary least squares (OLS) regressions, where the relationships between socioeconomic status and attitudes toward taxation will be analyzed.

In the first step of the analysis, a series of exploratory models is tested against the weighted data. Each model includes the two manifest tax variables—entered as nominal scale level variables in order to be able to distinguish nonlinear response patterns in the data—and country as a nominal-level covariate. While the effect of country is allowed to vary across clusters, all effects between the covariate "country" and the two manifest indicators are set to zero. This means: (a) that the characteristics of each cluster should be identical across

[13] The latent class analysis software used in this study is LatentGold 4.5.

countries, and (b) that cluster membership probabilities can vary across countries.

Moving to the independent variables that will be used in the analysis for testing H1, the main variable is socioeconomic status, which is an additive index standardized to vary between 0 and 100. It is based on three variables: Occupational Status, Education, and Family Income.

The first component in the socioeconomic index is the International Socio-Economic Index of Occupational Status (ISEI) (Ganzeboom, De Graaf, and Treiman 1992), which is based on ISCO88.[14] ISEI is rescaled in order to range between 0 and 100. The second component is education, where we have applied the standardized educational attainment level that is available in the ISSP background variable setup. This variable, containing six categories rescaled to range between 0 and 100, is derived from the national-specific variable asking for the respondents' highest completed level of education. The third component is equivalence income, which is family income weighted by household size, converted into percentiles. The weight equation used was:

Equivalence income = family income/ (household size * 0.73) (DAE 1998).

All three components have the same weight in the constructed index, which is scaled to run between 0 and 100. In cases where information was available for two components only, the index value was based on this information. We applied this strategy in order to decrease the number of missing values on the socioeconomic index.[15] To test whether the internal structure of the index is similar across countries, we conducted a factor analysis for each country. The results showed very little cross-country variation. In all countries, data formed a one-factor solution with similar component loadings (see Appendix Table A1).

The set of independent indicators selected to measure the size of the welfare state for testing H2, contains country-level data of "total tax revenue as percentage of GDP" and "total government spending as percentage of GDP", both collected from the Heritage Foundation online database. As noted above, studies within the Organization for Economic Cooperation and Development (OECD) area of countries show that both of these indicators are strongly correlated with the redistributive capacity of the welfare state.[16] For testing H3—on the role of economic development—we use one indicator: "GDP per capita" (2006 data), collected from the World Bank online database.

[14] Please note that the ISCO-based component of the index is a rather rough measure in the case of South Africa and Venezuela, since the ISCO coding is available only as a one-digit code.

[15] The total number of valid responses for the tax items is 42,162. Out of these, 40,125 have a valid response on the socioeconomic index.

[16] One caveat, however, is that data covering all 33 countries were available only from 2009. However, for the countries where time-series data are available, an analysis shows that the levels of government spending and taxation tend to be rather stable between 2006 and 2009.

2.4 Results

2.4.1 Attitudes toward Taxation in Thirty-Three Countries

Based on the results of the LCA model testing exercise explained before, we selected the five-cluster model as an adequate representation of the data (for details, see Appendix, Table A2). This means that five dominant types of tax preferences exist in the data. What characterizes each of these and how do they differ across countries in popularity? Let us first begin with the characteristics of each of the five tax preference types, as shown in Table 1.

The cell entries in Table 1 are item-response probabilities by cluster membership. For example, among those respondents classified into cluster 1, the probability to respond "for high-income earners … taxes are too low" is 99.8 percent, "taxes are about right" 0.1 percent, and "taxes are too high" 0.1 percent. The probability for cluster 1 respondents to respond "for low-income earners … taxes are too low" is 1.4 percent, "taxes are about right" 16.2 percent, and "taxes are too high" 82.4 percent. In other words, cluster 1 respondents prefer a more progressive distribution of taxes: taxes on low-income earners should be decreased, while taxes on high-income earners should be increased. Respondents in the second cluster also seem to prefer a more progressive distribution of taxes, particularly when it comes to tax cuts for low-income earners. These preference types will be labeled progressive1 (A) and progressive2 (B), respectively. The third cluster, content (C), comes closest to a preference pattern demonstrating contentment with current levels of taxation. The fourth group, called cut all (D), represents respondents favoring tax cuts for all income groups. Finally, the fifth cluster is composed of respondents favoring a higher tax burden for those with low incomes, while indicating that those with high incomes should become financially disburdened. This response pattern is hereafter called regressive (E).

Table 1. Attitudes towards taxation. Characteristics of five types of tax preferences (A-E)

	Cluster 1	***Cluster 2***	***Cluster 3***	***Cluster 4***	***Cluster 5***
	Progressive1 (A)	**Progressive2 (B)**	**Content (C)**	**Cut all (D)**	**Regressive (E)**
For high-income earners…					
Taxes are too high	0.1	23.2	15.4	87.0	58.3
Taxes are about right	0.1	36.5	75.8	12.9	28.9
Taxes are too low	99.8	40.4	8.9	0.1	12.9
For low-income earners…					
Taxes are too high	82.4	99.4	47.3	73.4	14.5
Taxes are about right	16.2	0.0	52.3	25.2	0.4
Taxes are too low	1.4	0.6	0.4	1.4	85.1

Note: Cell entries are response probabilities (%) by cluster membership. (n=42,162)

The empirical results are presented as follows. First, we will show the distribution of the five tax preferences per country. Thereafter follows the central empirical part of the study: the analysis of the relationship between socioeconomic status and attitudes toward redistributive taxation per country. Here, we have performed 33 single-country OLS-regressions for each of the 5 dependent variables (the 5 tax preferences) (in total 33*5=165 regressions). The dependent variables are cluster membership probabilities and each varies between 0 and 100. Apart from socioeconomic status, each regression contains 3 other independent variables: age, gender, and work status. These are used to control for compositional differences and are not of any analytical interest.

Based on the expected patterns according to the homo-economicus theory, a series of predictability measures (PM) is developed. The better the regression results meet the theoretical expectations, the higher the country's score on the predictability measures. These predictability measures are then used to test the extent to which the sizes of the predictability measures co-vary with the size of the welfare state and the level of economic development as specified in H2 and H3, respectively. Finally, we devote special attention to the countries displaying the most deviant attitude patterns. By checking question wording, translation issues, and sample coverage in these countries, the purpose is to analyze the extent to which these aberrant patterns may be caused by measurement errors related to the survey instrument, that is, data collection procedures.

Table 2: Tax Preference Cluster Sizes and Macro Indicators by Country

	Prog1 (A)	Prog2 (B)	Content (C)	Cut all (D)	Regressive (E)	(n) (A-E)	GDP/capita	Tax/ GDP	Government spending/GDP
Denmark	18.3	16.0	38.1	27.6	0.0	1,266	52,041	49.5	51.0
Finland	29.9	26.2	37.6	6.0	0.3	1,085	41,119	43.1	47.3
Norway	30.4	31.3	24.4	13.7	0.4	1,218	72,960	43.4	40.9
Sweden	20.7	42.9	20.1	16.3	0.1	1,097	46,256	48.9	52.5
France	42.0	3.4	29.5	20.5	4.7	1,564	36,545	45.0	52.3
Germany	52.8	15.5	25.7	5.5	0.5	1,434	36,400	40.8	44.2
Great Britain	19.9	16.0	42.9	19.5	1.7	830	42,448	37.9	44.0
Ireland	47.6	3.8	28.8	19.3	0.5	952	53,941	32.5	35.7
Netherlands	47.2	3.8	37.7	10.4	0.9	891	44,011	38.0	45.3
Switzerland	57.0	7.7	35.1	0.0	0.1	942	57,347	29.7	32.2
Croatia	19.9	71.6	8.3	0.0	0.2	940	11,360	23.4	42.0
Israel	23.0	29.3	26.7	18.9	2.1	1,177	21,582	37.0	46.4
Portugal	34.0	52.8	11.2	2.0	0.0	1,545	19,820	37.8	45.8
Spain	27.5	43.5	21.6	6.1	1.3	2,078	28,481	37.9	38.8
Czech Rep.	24.6	23.1	32.5	17.5	2.3	1,103	15,159	36.9	42.6
Hungary	60.9	3.7	24.3	4.1	7.0	898	11,343	39.9	49.7
Latvia	40.1	44.2	7.5	8.0	0.2	855	8,986	31.3	37.7
Poland	28.8	53.7	17.1	0.0	0.4	1,115	9,002	33.5	42.1
Slovenia	37.6	47.5	11.1	2.9	0.9	911	19,725	38.4	42.4
Russia	44.8	28.9	23.8	0.0	2.6	1,468	6,948	34.6	33.4
Australia	21.3	12.6	25.5	40.6	0.1	2,511	36,101	30.6	34.2
Canada	34.6	0.1	27.4	34.4	3.6	796	40,245	33.3	39.1
USA	49.5	0.0	33.6	11.5	5.5	1,379	46,437	28.3	37.4
New Zealand	9.2	6.9	34.8	48.6	0.5	1,099	26,336	36.0	40.3
Chile	14.2	63.3	14.7	2.6	5.2	1,331	9,371	18.9	18.6
Dominican Rep.	2.3	85.1	8.1	0.3	4.3	1,992	3,794	15.0	18.1
Uruguay	9.5	69.0	11.1	5.0	5.5	943	5,879	24.0	30.2
Japan	52.5	15.8	16.5	7.5	7.7	963	34,076	27.9	36.0
South Korea	79.7	0.4	12.5	5.5	2.0	1,541	20,917	28.7	28.9
Taiwan	61.7	9.3	26.7	2.0	0.3	1,630	30,084	13.7	17.8
Philippines	0.0	15.6	25.0	43.5	15.9	1,106	1,399	14.0	17.1
South Africa	11.8	0.0	32.3	42.1	13.8	2,469	5,660	26.6	27.8
Venezuela	9.3	27.0	31.1	11.9	20.8	1,033	6,748	17.0	35.7

Note: A-E= Cluster sizes. GDP/capita (in USD); Total tax revenue/GDP; Total government spending/GDP

In Table 2, the size of each of the 5 tax preference clusters (A–E) is shown by country. Speaking about broader patterns across 33 very heterogeneous countries is not an easy task. However, it becomes quite clear that in general, support for progressive taxation (A–B) is rather strong across countries, while support for regressive taxation (E) is generally low.

Focusing on the countries exhibiting the most deviant patterns compared to the majority of countries, we find the Philippines, South Africa, and Venezuela. In these countries, support for regressive taxation (E) is quite pronounced: 14 percent in South Africa, 16 percent in the Philippines, and even 21 percent in Venezuela. In addition, Table 2 shows that in both South Africa and the Philippines, the most common public preference is general tax cuts. Taken together, data suggest that these two citizenries express strong support for a prototypical neo-liberal right-wing tax system. Taking into account the great poverty among large parts of the population, the above reported results are, we believe, representative of empirical findings where researchers have reasons to question the quality of data. We will return to this issue later. In the next section, we analyze the relationships between socioeconomic status and attitudes toward redistributive taxation in order to assess the empirical support of H1–H3. When examining the results, we will also pay special attention to the response patterns observed for the above three countries.

2.4.2 Relationships between Socioeconomic Status and Attitudes toward Taxation in Thirty-Three Countries

In Table 3, columns A–E show the effect of socioeconomic status on each of the five dependent variables (standardized regression coefficients) per country. In order to assess the degree of correspondence between the assumptions formulated in H1—a negative relationship between socioeconomic status and support for redistributive taxation—we have constructed a series of predictability measures displayed in the last three columns of Table 3. These measures are based on the regression results presented in columns A–E. Below follows the theoretical rationale for the design of the predictability measures, as well as how they are constructed.

According to the basic homo-economicus theory, those economically less well-off should score high on the progressive taxation preferences (A, B) and low on preferences for general tax cuts (D) as well as on support for regressive taxation (E). For contented respondents (C) it is not fruitful to specify an expected direction of the relationship. In some countries, the “leftist” position may be to defend the current levels of taxation against those arguing for lowered tax rates; whereas in other countries, the dividing line may go between those in favor of increased taxation and those content with taxes as they are. Therefore, the actual sign of the contented

preference coefficient is of no interest.

The first *predictability measure (PM I),* measuring the strength of the association between socio-economic status and preferences for redistributive taxation is therefore specified as:

$$[\ PM\ I = (-\ beta_a - beta_b + |beta_c| + beta_d + beta_e)*100\ ]$$

However, in order to check the robustness of the results we have designed two complementary measures. The second predictability measure, *PM II,* includes only significant beta coefficients (10% level) and for *PM III,* only the three strongest beta coefficients are included. The theoretical rationale underlying *PM III* is that the number of dominant lines of political conflict over tax policy in most countries is likely to be limited.

Columns A and B in Table 3 display the associations between socioeconomic position and attitudes toward progressive taxation. While it is theoretically expected that both of these relationships should be negative, we find that this "twin pattern" occurs only in 19 out of 33 countries, and in some cases only one relationship is significant. It is, however, worth underlining that virtually all of the North and West European countries display this twin pattern (except Great Britain). In the remaining 14 countries, we find either combinations of positive and negative associations with progressive taxation, or—in a small number of countries: the Philippines, Venezuela, Uruguay, and Hungary—no significant relationships at all.

When it comes to the relationship between socioeconomic position and the likelihood of preferring a smaller welfare state by reducing taxes—and thereby the decreasing redistributive capacity of the state—in a majority of countries the expected positive relationship cannot be observed (column D). In 16 countries, the relationship is nonsignificant and in 4 countries it is significantly negative, that is, the support for general tax cuts comes disproportionally from the lower strata on the socioeconomic ladder. In thirteen countries, however, we can observe the expected significant positive relationship.

In column C, the focus is on those who are content with the income tax. We can observe a predominantly positive relationship between socioeconomic position and tax contentment in most of the countries, although the magnitude of the relationship varies substantially. Most of the pronounced relationships pertain to relatively affluent and advanced welfare states.

Table 3: Attitudes towards taxation by socio-economic status in thirty-three countries

	A	B	C	D	E	PM I*	PM II**	PM III***
	Prog1	Prog2	Content	Cut all	Regressive			
Expected direction	-	-	||	+	+			
Denmark	**-17.0**	**-21.2**	3.9	**19.8**	0.5	62.4	58.0	58.0
Finland	**-25.3**	**-19.3**	**26.9**	**23.2**	-1.7	93.0	94.7	75.4
Norway	**-10.2**	**-14.7**	**19.1**	1.9	**7.1**	52.9	51.1	44.0
Sweden	**-27.5**	**-13.9**	**12.9**	**31.0**	-2.6	82.8	85.4	72.4
France	**-11.0**	**-19.5**	**6.0**	2.1	**16.1**	54.6	52.5	46.5
Germany	**-10.6**	**-16.4**	**20.6**	1.3	-2.2	46.8	47.7	47.7
Great Britain	0.5	**-19.1**	**8.7**	-4.1	1.7	24.9	27.9	23.7
Ireland	-2.1	**-13.2**	**13.6**	**-11.0**	5.3	23.2	15.8	15.8
Netherlands	**-25.4**	**-19.4**	**20.3**	**18.7**	**7.6**	91.4	91.4	65.1
Switzerland	**-11.9**	-4.4	**13.4**	2.9	**6.6**	39.2	31.9	31.9
Croatia	1.4	**-6.5**	**6.3**	1.7	-4.7	8.4	12.8	8.1
Israel	**-9.9**	**-16.0**	**11.3**	**7.6**	**6.6**	51.3	51.3	37.2
Portugal	-3.0	**-7.6**	**12.6**	0.7	1.2	25.1	20.2	23.2
Spain	**-5.3**	**-9.7**	**15.0**	-0.3	-3.5	26.2	29.9	29.9
Czech Rep.	**-22.5**	**-10.4**	**14.8**	**16.1**	2.9	66.5	63.7	53.3
Hungary	0.0	3.0	-2.0	2.5	1.2	2.7	0.0	1.5
Latvia	1.2	**-7.1**	3.4	3.9	-0.3	12.8	7.1	14.4
Poland	**-20.1**	**6.3**	**14.2**	**15.3**	1.0	44.3	43.2	49.5
Slovenia	**-7.9**	**-7.1**	**15.8**	**17.9**	**-7.4**	41.3	41.3	41.7
Russia	4.4	**-8.7**	1.8	-3.1	-1.8	1.2	8.7	1.2
Australia	0.1	**-12.4**	**11.4**	**-5.6**	1.6	19.6	18.2	18.2
Canada	**-11.4**	**-14.7**	6.0	**6.5**	4.4	43.1	32.7	32.7
USA	0.5	**-21.0**	2.8	**-5.1**	-0.7	17.5	15.9	18.7
New Zealand	**-11.8**	**-17.4**	0.6	**9.9**	0.4	40.0	39.0	39.0
Chile	4.5	**-11.0**	**13.7**	3.0	**-6.1**	17.1	18.6	18.6
Dominican Rep.	**5.0**	-3.2	-0.1	0.9	2.8	2.0	-5.0	1.0
Uruguay	0.5	4.6	-0.7	**6.7**	**-9.2**	-6.9	-2.5	-7.1
Japan	-5.5	**-7.3**	3.0	**6.2**	**6.7**	28.7	20.2	20.2
South Korea	**-8.3**	-2.2	4.1	**9.1**	4.2	27.9	17.4	21.6
Taiwan	**12.3**	**-12.9**	**-9.6**	**-8.0**	1.9	4.2	2.2	10.2
Philippines	1.8	4.9	1.7	4.5	**-9.4**	-9.9	-9.4	-9.9
South Africa	**-4.1**	**-9.4**	**7.5**	1.5	**-6.3**	16.2	14.7	10.6
Venezuela	1.3	3.4	**8.5**	-2.6	**-8.8**	-7.7	-0.4	-3.8

Notes: Cell entries in columns A–E are standardized regression coefficients *100. PM I–III are predictability measures. PM I–III: * = All five beta coefficients included; ** = only significant (10 percent level) beta coefficients are included; *** = only the three strongest beta coefficients are included. Bold coefficients are significant (10 percent level). Age, gender, and work status are included as independent variables but not shown. Numbers in columns A–E are rounded. N=42,125.

Finally, as shown previously in Table 2, the proportion of people supporting a more regressive distribution of taxes is quite small in most countries, which to some extent may explain why the association between socioeconomic position and attitudes toward regressive taxation is nonsignificant in a large number of countries (Table 3, column E). Two findings are worth emphasizing. First, the expected positive relationship is manifested in six countries only. Second, in the three countries where regressive tax preferences have shown to be particularly common—the Philippines, Venezuela, and South Africa—we observe a negative significant relationship: the lower the socioeconomic position the higher the support for a more regressive distribution of taxes.

The empirical results in Table 3 demonstrate that the similarity between the theoretically expected relationship and empirical observations varies extensively across countries. We will now examine the extent to which the size of the welfare state and the level of economic development "explain" this variation.

In Table 4, the correlation coefficients (Pearson's r) between the two indicators of the size of the welfare state, as well as the indicator of economic development, and each of the three predictability measures are displayed. The associations are quite strong and positive indicating that the strength of the expected relationships between socioeconomic position and attitudes toward redistributive taxation co-vary with the size of the welfare state and the level of economic development.

Table 4: Associations between Predictability Measures (PM I-III) and Macro Indicators

	PM I	PM II	PM III
Total tax revenue/GDP	74	75	76
Total government spending/GDP	65	67	66
GDP/capita	58	54	57

Notes: Cell entries are correlation coefficients*100 (Pearson's r)(n=33); All correlations are significant at the 1 percent level.

The results indicate that the expected relationships between socioeconomic position and attitudes toward redistributive taxation are far more often met for the Western, especially the North European countries, than for the less affluent non-Western countries. The predictability score is substantially lower for this latter group of countries. Uruguay, Venezuela, and the Philippines even report a negative score on all three variants of the predictability measure.

Overall, the results suggest that the fit between theory and empirics is much better in affluent Western compared to less affluent non-Western countries. Moreover, within the latter group, the structure of the observed patterns shows a much larger cross-country variation. Arguably,

some of the results cause suspicions about the quality of data. This relates in particular to the empirical results for the Philippines, Venezuela, and South Africa. In the next section we take a closer look at these three countries by examining issues related to translation and sample coverage.

2.4.3 Venezuela, South Africa, and the Philippines: Measurement Errors Related to Translation or Sample Coverage?

In this section we focus on the three countries in which preferences toward taxation are strikingly different from those in the other examined countries. In these countries, many respondents believe that the poor should carry a higher share of the tax burden, and that taxes on those with higher incomes should be lowered. Even more peculiar, the data reveal that a surprisingly large share of respondents embracing these views does not belong to the upper classes of society. On the contrary, they are, in fact, rather poor themselves. First, we examine potential errors related to translation. Second, we examine sample coverage focusing on the representation of the poor in the samples.

The translation process of converting the British English ISSP source questionnaire into native languages can produce errors (Harkness, Villar, and Edwards 2010b; Smith 2011; Weisberg 2005; Willis et al. 2010). The used tax items use a five-point scale ranging from [taxes are] "much too high" to "much too low." Although the ISSP Working Principles (ISSP Research Group 2012) state that primary researchers in all participating countries must oversee an appropriate translation into the national language/languages, a simple translation error reversing response categories could produce the observed patterns.[17]

This seems a plausible source of error especially for the Philippines and South Africa because in both countries the ISSP 2006 survey was conducted in six different languages, with uneven numbers of respondents using the various questionnaires.[18] Reversing the direction of the response scale in a dominant language could be enough to cause outcomes that look like high levels of support for regressive taxation. Moreover, since it is likely that many of those involved in the data processing do not understand all of these languages, possible mistakes may not be easily discovered.

[17] ISSP Working Principles: "non-literal translations, that is, 'culturally equivalent questions,' are to be notified to the group and indicated in the documentation." Apart from that there are no mandatory rules for the procedure of the translation process.

[18] The languages covered by the Philippine survey are English, Tagalog, Ilocano, Bicolano, Ilonggo, and Cebuano. In South Africa the languages are English, Afrikaans, Venda, Xhosa, Zulu, and Tsonga. According to the ISSP Study Monitoring Report (Scholz et al. 2008), all translations were carried out by trained translators and checked by language experts.

For South Africa and Venezuela this potential error can be excluded on the basis of language checks conducted by native speakers.[19] Thus, the response patterns do not seem to be caused by errors in either translation or wording in these two countries. For examining the Philippines data we obtained a variable from the national data set indicating the language in which the interview was conducted.[20] An analysis of the response patterns across language groups shows that three of five language groups have a somewhat larger probability of holding nonredistributive tax preferences. However, the probability that a translation error would appear in three languages simultaneously seems unlikely. Besides, if such a massive translation error were true, the support for nonredistributive taxation would most certainly be much stronger than it actually is in the Philippines. In our judgment, the results for Philippines do not seem to depend on translation errors.

Another source of error worth considering is the respondent selection issue, in this case a sample coverage error. Such an error could be responsible for the data outcome (Heeringa and O'Muircheartaigh 2010). Coverage errors generate bias due to the omission of non-covered sample units (Weisberg 2005). The Philippines, South Africa, and Venezuela have large areas of informal settlements with significant numbers of poor people making a living outside the formal economy (Perazzi, Merli, and Paredes 2010; Schneider 2002; Wills 2009). It is easy to imagine that sampling procedures and/or interview situations might be difficult in such areas. If these areas are not covered in the sample, the number of respondents belonging to the poorest classes of society would be severely underestimated; the very same classes that have the strongest reasons for supporting a more redistributive tax policy.

Moving beyond the obligatory methodological report on interview situations in informal settlements, we queried the primary researchers about this issue. We received responses from the Philippines and Venezuela and the researchers assured us that these areas had been covered. Interviewers in the Philippines were given special training on area coverage, and coverage of slum areas was said to not pose much of a problem. In Venezuela a number of people in poor districts were employed as field staff, which facilitated access to the dangerous zones as they lived there themselves and knew how to operate in these areas.

The sample coverage issue was also approached from another angle, using national statistics in order to assess the percentage of the population living below the subsistence level.[21] For each

[19] Here we would like to thank Ivet Solanes Ros and Webster Whande for their support.
[20] Thanks to Gerado A. Sandoval from the Social Weather Stations in the Philippines.
[21] The poverty and subsistence thresholds, respectively, are absolute figures, comprising people who cannot afford to buy a national specific minimum basket of goods. The subsistence threshold lies even below the poverty line and indicates the very minimum level of what a person needs to survive. The definitions of these levels are set by the countries.

country, the obtained estimate was thereafter compared with the income distribution in the country sample.[22] The national specific family income variable and household size were used to determine the equivalence income.[23]

Beginning with the Philippines, official statistics estimate the subsistence threshold line in 2006 to be PHP835 per month, and 15 percent of the general population is below this line.[24] In the ISSP sample, we find that 12.5 percent of the sample is below the line. Thus, differences between the two estimates are quite small or at least not large enough to explain the high support for regressive taxation in the Philippines.

For South Africa, we can only apply a quite rough measure of the poverty rate. First, the income information in the South African ISSP sample is available only in the form of income classes. The official subsistence threshold line identifies a specific income value. Second, the ISSP income variable has a high nonresponse rate (27 percent), thus making it difficult to judge the extent to which the sample is representative of the population in terms of income distribution.

Various possible poverty measures for South Africa can be found in the literature. For our purposes, the most suitable measure found was "the necessary amount of money to purchase enough food to meet the basic daily food-energy requirements for an average person in South Africa over one month." In 2005, it was estimated to be ZAR250.[25] Based on these facts, we decided to use the lowest income class available in the ISSP data (ZAR250 per month) as the threshold. Although prerequisites are not ideal, the analysis should reveal whether the poor are severely underrepresented. According to official statistics, the poverty rate in 2006 was 23 percent. Our estimate on the ISSP sample shows a poverty rate close to 25 percent. The ISSP sample thus comes fairly close to representing the poor population of South Africa.

[22] We are aware that there are additional dimensions of poverty that cannot be comprehended by measuring pure income levels. Furthermore, there exist different income-based measures of poverty as well as poverty lines (see, e.g., Ravallion 2010). Apart from that, in urban economies with large informal sectors, income flows may be erratic. The same is true for agrarian economies where household incomes rely on the harvest cycle (Coudouel, Hentschel, and Wodon 2002). Nevertheless, the comparison of income statistics with the income information in the survey data should be sufficient to figure out whether an entire group of the society might be excluded from the survey.

[23] Incomes tend to be underestimated in surveys, especially with only one income question, because respondents simply forget to include all income sources (DAE 1998). However, that does not account for those respondents living at the existence minimum. Those who have very little, are usually able to keep track of it. Since the age structure of the household members is unknown, the following formula was used: equivalence income ¼ family income / (household size * 0.73). (For further information regarding the construction of equivalence income see DAE [1998].)

[24] National Statistical Coordination Board, Philippines, 2006.

[25] The Presidency of South Africa (2008) Development Indicators 2008: Van der Berg (2007) based on AMPS of various years (1993–2007), Bhorat (2007) based on Statistics SA IES data (1995, 2000, and 2005), and Statistics South Africa and National Treasury (2007:23). ZAR stands for South African Rand.

In Venezuela, INE (Instituto Nacional De Estadística) data differentiate between poor and extremely poor people. As done for the Philippines and South Africa we focus on the group of extremely poor. In Venezuela the percentage of the population belonging to this category is calculated on the basis of a basket of goods. Since these data do not contain any specific monetary value, a direct comparison of official statistics and ISSP data is not possible. Nevertheless, to get an idea of the proportions, we applied the World Bank's international poverty line of USD 1.25 per day as a rough estimate. Based on this estimate, the ISSP data underestimate the poverty rate for Venezuela, by four percentage units (8 percent compared to 12 percent). This result also does not suggest that the collected data severely misrepresent the actual socioeconomic structure in Venezuela.

In sum, it seems that the observed aberrant response patterns in Venezuela, South Africa, and the Philippines are not related to obvious translation errors, such as reversed response scales, or to sample coverage.[26]

2.5 Conclusion

Beginning with a simple rational choice-inspired homo-economicus hypothesis saying that those who are better off in the socioeconomic hierarchy should have less reason to support state-organized economic redistribution compared to those situated at the lower levels in the socioeconomic hierarchy, this article set out to empirically study the relationships between socioeconomic status and attitudes toward redistributive taxation across 33 countries.

The empirical results demonstrated substantial cross-country variation regarding the correspondence between empirical observations and theoretical expectations. In some countries, the relationships between socioeconomic position and attitudes toward redistributive taxation were clearly in line with expectations. However, in many countries, relationships were weak and in some countries, even inverted.

When faced with such tremendous cross-national variation in response patterns, a common strategy among researchers is to question the quality of the data and data collection procedures for the countries deviating strongly from theoretical expectations. Somewhat bluntly put, this "questioning" can be expressed in two different ideal-typical strategies: implicit questioning (avoiding use of the data) or explicit questioning (using the data, and

[26] Further analyses have been done taking into account different regions and therewith different regional subsistence lines for the Philippines and South Africa. The results do not change the cross-regional data outcomes.

basically arguing that it is bad data). The strategy chosen in this study was very different and quite uncommon in cross-national attitude research. Our main argument was that cross-national variation in attitudes as a phenomenon may point to theory limitations in terms of generalization, that is, that the chosen theory may be more valid in some countries than in others. In other words, the lack of fit between theory and empirical observations may be rooted in problems related to theory rather than the quality of data collection procedures.

Building on insights and results from previous research we formulated two "cultural distance" hypotheses, both arguing that the correspondence between the homo-economicus theory and empirics should indeed vary systematically across countries. The first one focused on the role of the welfare state and institutional feedback effects on public opinion and suggested that the more encompassing the welfare state, the stronger the correlation between socioeconomic status and redistributive taxation. The second hypothesis, related more to the level of economic affluence and associated scientific dominance across countries, stated that the higher the level of economic development in a country, the stronger the correlation between socioeconomic status and redistributive taxation. Both of these hypotheses received considerable empirical support. The relationship between socioeconomic status and support for redistributive taxation was found to be substantially stronger in the wealthy Western welfare states—particularly among those of Northern Europe—than in the poor non-Western countries lacking any institutional design reminiscent of a welfare state.

For some countries—the Philippines, Venezuela, and South Africa—the empirical results deviated from theoretical expectations to such an extent that we felt compelled to perform in-depth analyses of some key aspects of the data collection procedures in these countries regarding translation and sample coverage. However, we did not find any clear evidence that the observed results were rooted in measurement errors related to either translation or sample coverage. Moreover, when keeping in mind that the apparent misfit between the homo-economicus theory and empirics relates to virtually all of the rather poor non-Western countries, explanations related to measurement error in the data collection procedures used in the Philippines, South Africa, and Venezuela seem even less plausible.

To conclude, we believe that the results in this study show that socioeconomic indicators such as income, education, and occupational status are relevant factors explaining public support for state-organized economic redistribution in many of the Western countries, but not in the poor non-Western countries. What are the broader implications of this finding? In the context of this study, we would like to stress that theoretically deviant empirical observations in cross-national research may not be related to bad data quality, but to theory generalization

limitations. The weak correspondence between the homo-economicus theory and empirics observed in most of the non-Western countries may indicate that Western theoretical assumptions are simply not valid.

2.6 References

Åberg, Rune. 1989. “Distributive Mechanism of the Welfare State.” *European Sociological Review* 5:167–182.

Blasius, Jörg, and Victor Thiessen. 2006. “Assessing Data Quality and Construct Comparability in Cross-National Surveys.” *European Sociological Review* 22:229–242.

Blasius, Jörg, and Victor Thiessen. 2012. *Assessing the Quality of Survey Data*. London: Sage.

Corneo, Giacomo, and Hans-Peter Grüner. 2002. “Individual Preferences for Political Distribution.” *Journal of Public Economics* 83(1):83–107.

Coudouel, Aline, Jesko S. Hentschel, and Quentin T. Wodon. 2002. “Poverty Measurement and Analysis.” Pp. 29–74 in *The PRSP Sourcebook.* Washington, DC: World Bank.

Deutsche Arbeitsgemeinschaft Epidemiologie (DAE). 1998. “Messung und Quantifizierung soziodemographischer Merkmale in epidemiologischen Studien [Measurement and quantification of sociodemographic indicators in epidemiological studies].” Pp. 7–38 in *Messung soziodemographischer Merkmale in der Epidemiologie [The measurement of sociodemographic indicators in epidemiology]*, edited by Wolfgang Ahrens, Bärbel-Maria Bellach, and Karl-Heinz Jöckel. Munich: MMV Medizin Verlag München.

Edlund, Jonas. 1999. “Attitudes to Income Redistribution and Taxation in Social Democratic and Liberal Welfare State Regimes: A Comparison Between Sweden, Great Britain, and the United States.” Pp. 13–110 in *Modern Society and Values: A Comparative Analysis Based on the ISSP Project*, edited by Niko Tos, Peter P. Mohler, and Brina Malnar. Ljubljana: University of Ljubljana and ZUMA.

Edlund, Jonas. 2007. “Class Conflicts and Institutional Feedback Effects in Liberal and Social Democratic Welfare Regimes: Attitudes towards State Redistribution and Welfare Policy in Six western Countries.” Pp. 30–79 in *The Political Sociology of the Welfare State: Institutions, Cleavages, Orientation*, edited by Stefan Svallfors. Stanford, CA: Stanford University Press.

Edlund, Jonas, and Rune Åberg. 2002. “Social Norms and Tax Compliance.” *Swedish Economic Policy Review* 9(1):201–228.

Edlund, Jonas, and Arvid Lindh. 2015. The Democratic Class Struggle Revisited: The Welfare State, Social Cohesion and Political Conflict.” *Acta Sociologica* 58(4):311-328.

Edlund, Jonas, and Ingemar J. Sevä. 2013. “Exploring the ‘Something for Nothing’ Syndrome: Confused Citizens or Free Riders? Evidence from Sweden.” *Scandinavian Political Studies* 36:293–319.

Esping-Andersen, Gøsta. 1992. “The Three Political Economies of the Welfare State.” Pp. 92–123 in *The Study of Welfare State Regimes*, edited by J.E. Kolberg. Armonk, NY: M.E. Sharpe.

Ferrera, Maurizio. 1996. “Southern Model of Welfare in Social Europe.” *Journal of European Social Policy* 6(1):17–37.

Ganzeboom, Harry B. G., Paul M. De Graaf, and Donald J. Treiman. 1992. “A Standard International Socio-Economic Index of Occupational Status.” *Social Science Research* 21(1):1–56.

Hagenaars, Jacques A., and Loek C. Halman. 1989. “Searching for Ideal Types: The Potentialities of Latent Class Analysis.” *European Sociological Review* 5:81–96.

Harkness, Janet A., Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith. 2010a. Survey Methods in *Multinational, Multiregional and Multicultural Contexts*. New York: Wiley.

Harkness, Janet A., Ana Villar, and Brad Edwards. 2010b. “Translation, Adaptation and Design.” Pp. 117–141 in *Survey Methods in Multinational, Multiregional and Multicultural Contexts*, edited by Harkness et al. New York: Wiley.

Heeringa, Steven G., and Colm O’Muircheartaigh. 2010. “Sampling Designs for Cross-Cultural and Cross-National Survey Programs.” Pp. 251–268 in *Survey Methods in Multinational, Multiregional and Multicultural Contexts*, edited by Harkness et al. New York: Wiley.

Heritage Foundation. 2009. Online database. Retrieved June 15, 2015 (http://www.heritage.org/index/explore).

Instituto Nacional de Estadística (INE), República Bolivariana de Venezuela. 2006. Retrieved March 14, 2012 (http://www.ine.gov.ve/ index.php?option=com_content&view= category&id=104:pobreza&Itemid=45).

International Social Survey Programme (ISSP) Research Group. 2012. *ISSP Working Principles*. Retrieved March 1, 2015 (http://www.issp.org/page.php?pageId=170).

Kankaras, Miloš, Guy Moors, and Jeroen K. Vermunt. 2011. “Testing for Measurement Invariance with Latent Class Analysis.” Pp. 359–384 in *Cross-Cultural Analysis: Methods and Applications*, edited by E. Davidov, P. Schmidt, and J. Billiet. New York: Routledge.

Korpi, Walter. 2006. “Power Resources and Employer-Centered Approaches in Explanations of Welfare States and Varieties of Capitalism: Protagonists, Consenters, and Antagonists.” *World Politics* 58:167–206.

Korpi, Walter, and Joakim Palme. 1998. “The Paradox of Redistribution and Strategies of Equality: Welfare State Institutions, Inequality and Poverty in the Western Countries.” *American Sociological Review* 63:661–687.

Korpi, Walter, and Joakim Palme. 2003. “New Politics and Class Politics in the Context of Austerity and Globalizsation: Welfare State Regress in 18 Countries, 1975–95.” *American Political Science Review* 97:425–446.

Kurzman, Charles. 2014. “World Values Lost in Translation.” Retrieved April 1, 2015 (http://www.washingtonpost.com/blogs/monkey-cage/wp/2014/09/02/world-values-lost-in-translation/).

Magidson, Jay, and Jeroen K. Vermunt. 2003. “Latent Class Models.” In *Handbook for Quantitative Methodology*, edited by David Kaplan. Newbury Park, CA: Sage.

Mellon, Jonathon. 2011. “Examining Survey Translation Validity Using Corpus Linguistics.” Working paper in the Nuffield Working Papers Series in Politics.

Mettler, Suzanne, and Joe Soss. 2004. “The Consequences of Public Policy for Democratic Citizenship: Bridging Policy Studies and Mass Politics.” *Perspectives on Politics* 2:55–73.

National Statistical Coordination Board (NSCB), Philippines. 2006. Annual per Capita Poverty Thresholds, Poverty Incidence and Magnitude of Poor Families: 2000, 2003 and 2006. Retrieved November 14, 2012 (http://www.nscb.gov.ph/poverty/2006_05mar08/table_1.asp).

Organization for Economic Cooperation and Development (OECD). 1990. *The Personal Income Tax Base*. Paris.

Öun, Ida. 2012. *Conflict and Concord in Work and Family. Family Policies and Individual’s Subjective Experiences.* Doctoral dissertation. Department of Sociology, Umeå Universitet.

Perazzi, Josefa R., Giampaolo O. Merli, and Laura C. Paredes. 2010. “The Size of the Informal Economy in Venezuela.” *El Norte-Finnish Journal of Latin American Studies* 5:1–12.

Persson, Anna, and Bo Rothstein. 2015. “It’s My Money: Why Big Government May Be Good Government.” *Comparative Politics* 47(2):231–249.

Presidency of South Africa. 2008. Development Indicators 2008. Retrieved December 1, 2014 (http://www.info.gov.za/ view/DownloadFileAction?id=85063).

Ravallion, Martin. 2010. “Poverty Lines across the World.” World Bank Policy Research Working Paper 5284. Washington, DC. Retrieved March 1, 2015 (http://www.wds.worldbank.org/external/default/WDSContentServer/ IW3P/IB/2010 /04/27/000158349_20100427143536/Rendered/INDEX/WPS5284.txt).

Rothstein, Bo. 2000. “Trust, Social Dilemmas and Collective Memory.” *Journal of Theoretical Politics* 12:477–501.

Rothstein, Bo, and Sven Steinmo. 2002. (eds.). *Restructuring the Welfare State: Political Institutions and Policy Change.* New York: Palgrave.

Schneider, Friedrich. 2002. *Size and Measurement of the Informal Economy in 110 Countries around the World.* Washington, DC: Rapid Response Unit, World Bank.

Scholz, Evi, Timo Faaß, Janet Harkness, and Marleen Heller. 2008. *ISSP Study Monitoring 2006. Report to the ISSP General Assembly on Monitoring Work Undertaken for the ISSP* by Gesis-ZUMA, Gesis Methodenberichte 11. Mannheim: Gesis. Retrieved March 1, 2015 (http://www.gesis.org/issp/issp-modules-profiles/role-of-government/ 2006/).

Smith, Tom W. 2011. “Refining the Total Survey Error Perspective.” *International Journal of Public Opinion Research* 23(4):464–484.

Soss, Joe, and Sanford F. Schram. 2007. “A Public Transformed? Welfare Reform as Policy Feedback.” *American Political Science Review* 101(1):111–127.

Statistics South Africa and National Treasury. 2007. A National Poverty Line for South Africa. Retrieved March 1, 2015 (http://www.treasury.gov.za/publications/other/povertyline /Treasury%20StatsSA%20poverty%20line%20 discussion%20paper.pdf)

Steinmo, Sven, Kathleen Thelen, and Frank Longstreth, eds. 1992. *Structuring Politics: Historical Institutionalism in Comparative Analysis.* Cambridge: Cambridge University Press.

Svallfors, Stefan. 2006. *The Moral Economy of Class.* Stanford, CA: Stanford University Press.

Svallfors, Stefan. 2007. “Introductory Chapter.” Pp. 1–29 in *The Political Sociology of the Welfare State: Institutions, Social Cleavages and Orientations*, edited by Stefan Svallfors. Stanford, CA: Stanford University Press.

Wallerstein, Immanuel. 1997. “Eurocentrism and Its Avatars: The Dilemmas of Social Science.” *New Left Review* 226:93–107.

Weisberg, Herbert F. 2005. *The Total Survey Error Approach: A Guide to the New Science of Survey Research.* Chicago: University of Chicago Press.

Wilensky, Harold L.. 2002. *Rich Democracies: Political Economy, Public Policy, and Performance.* Los Angeles, CA: University of California Press.

Willis, Gordon B., Martha Stapleton Kudela, Kerry Levin, et al. 2010. "Evaluation of a Multistep Survey Translation Process." Pp. 141–156 in *Survey Methods in Multinational, Multiregional and Multicultural Contexts,* edited by Harkness et al. New York: Wiley.

Wills, Gabrielle. 2009. "South Africa's Informal Economy: A Statistical Profile." Wiego Working paper (Urban Policies) 6. Retrieved March 1, 2015 (http://wiego.org/sites/wiego.org/files/publications/files/Wills_WIEGO_ WP6.pdf).

World Bank. 2006. Tax revenue (% GDP). Retrieved November 14, 2014 (http://data.worldbank.org/indicator/GC.TAX.TOTL.GD.ZS?page=1&order=wbapi_data_value_2007%20wbapi_data_value%20wbapi_data_value-first& sort=asc).

Appendix

Table A1: Socioeconomic Index: Principal Component Factor Analysis Using Varimax Rotation

Country	Education	Family Income	ISEI	Variance in %
Denmark	0.79	0.70	0.85	61.1
Finland	0.84	0.68	0.85	62.8
Norway	0.77	0.73	0.84	61.1
Sweden	0.79	0.70	0.83	60.0
France	0.79	0.78	0.84	64.4
Germany	0.86	0.70	0.87	66.2
Great Britain	0.79	0.77	0.83	63.3
Ireland	0.84	0.70	0.84	63.6
Netherlands	0.77	0.72	0.83	59.6
Switzerland	0.87	0.63	0.88	64.2
Croatia	0.87	0.74	0.88	69.2
Israel	0.83	0.59	0.85	58.9
Portugal	0.84	0.72	0.86	65.2
Spain	0.83	0.79	0.86	68.7
Czech Rep.	0.89	0.61	0.89	65.3
Hungary	0.87	0.63	0.87	63.6
Latvia	0.82	0.62	0.88	60.9
Poland	0.88	0.73	0.88	69.1
Slovenia	0.87	0.81	0.90	74.2
Russia	0.79	0.62	0.84	57.2
Australia	0.75	0.71	0.78	55.9
Canada	0.71	0.61	0.78	49.5
USA	0.79	0.78	0.82	63.5
New Zealand	0.73	0.67	0.78	53.1
Chile	0.83	0.79	0.86	68.1
Dominican Rep.	0.80	0.72	0.84	62.0
Uruguay	0.86	0.81	0.89	72.8
Japan	0.78	0.73	0.80	58.9
South Korea	0.82	0.79	0.86	67.4
Taiwan	0.82	0.77	0.85	66.7
Philippines	0.78	0.71	0.84	60.1
South Africa	0.85	0.76	0.85	67.5
Venezuela	0.82	0.77	0.68	57.7

Table A2: Model fit statistics (n=42,162)

Model	L^2	CAIC (L^2)	df	L^2 reduction %
1	12679.4	9650.6	260	0.0
2	5536.3	2938.5	223	56.3
3	3316.1	1149.3	186	73.8
4	1834.7	99.0	149	85.5
5	877.7	-427.1	112	93.1
6	452.1	-421.6	75	96.4

Notes: The L^2 value of model 1 (the 1-cluster model), the baseline model, indicates the maximum association between the manifest variables that can be explained by any latent class variable. Judging by the consistent Akaike information criterion statistic (CAIC) (the lower the value, the better the model), the baseline model should be rejected in favor of a more complex model (Kankaras, Moors, and Vermunt 2011). By relying on the CAIC statistic, it is suggested that the five-cluster model should be chosen (model 5). The L^2 value is reduced by 93.1 percent. Furthermore, and very important for a good model fit, the three error correlations (i.e., bivariate residuals that the latent variable does not account for) between the manifest variables are all nonsignificant (table available on request).

2017-11-13
Jonas Edlund
Professor of Sociology
Umeå University, Sweden

UMEÅ UNIVERSITET

The journal article "Observing Unexpected Patterns in Cross-National Research: *Blame Data, Theory, or Both? Attitudes toward Redistributive Taxation in Thirty-Three Countries"* (International Journal of Sociology, 45:4, 327-347) was authored by Insa Bechert (GESIS Leibniz Institute for the Social Sciences) and Prof. Jonas Edlund (Umeå University).

The work on this article was shared as follows:

Task	Contributor	Contribution in %
Development of key ideas	Insa Bechert Jonas Edlund	50 50
Development of the experimental design	Insa Bechert Jonas Edlund	50 50
Literature review	Insa Bechert	80
Provision of the theoretical background on the role of welfare states/economic development for attitudes towards redistribution	Jonas Edlund	100
Data preparation	Insa Bechert	80
Construction of the socio-economic index	Insa Bechert	100
Data quality checks (translation/sample coverage)	Insa Bechert	100
Statistical analyses	Insa Bechert Jonas Edlund	50 50
Interpretation of the results	Insa Bechert Jonas Edlund	50 50

Umeå, 17-11-13 [redacted]

Prof. Jonas Edlund

Cologne, Nov 29th, 2017 [redacted]

Insa Bechert

This is a pre-print of an article published in *Zeitschrift für Religion, Gesellschaft und Politik*. The final publication is available at Springer via: https://doi.org/10.1007/s41682-018-0016-z.

3. Comparing Religiosity Cross-Nationally - *About Invariance and the Role of Denomination*[27]

Abstract

For quite some time there has been widespread consensus in the social sciences that religiosity is a multidimensional and multifaceted phenomenon (see for example Glock 1962; Storm 2009). Individuals may be very religious with respect to one dimension and less so towards another. Previous research has identified various typologies, also called religious profiles, showing different combinations of religious multidimensionality within and across countries. This paper identifies dominant cross-national profiles of religiosity and (1) examines whether there is one valid typology worldwide, or if some countries show profiles more similar to one another than to others; (2) tests the results for invariance to examine whether it is actually appropriate to compare the empirically found typologies of religiosity cross-nationally and (3) looks at the impact of denomination for profile formation. The results reveal that (1) there is a valid typology worldwide, but instead of showing specific characteristics, profiles can be ordered on a single latent continuum from low to high levels of religiosity. (2) It is almost impossible to establish full invariance across countries with a comprehensive set of variables measuring religiosity, but partial homogeneity can be achieved. Contrary to all theoretical expectations, the analyses imply (3) that denomination is not a crucial impact factor for religious profile formation.

Keywords: religiosity, LCA, invariance

[27] Published in *Zeitschrift für Religion, Gesellschaft und Politik*

3.1 Introduction

Even if God was already declared dead by 19th century European philosophers, ongoing global debates on secularization and recent religion-based conflicts suggest that studies on religion will only gain in importance in the foreseeable future. Knowledge about the nature of individual religiosity enables examinations of increases and declines in religiosity in societies as it sheds light on mechanisms that might trigger conflicts.

Empirical research has shown that religiosity is a multifaceted phenomenon (Glock 1962; Cornwall et al. 1986; King and Hunt 1975; Storm 2009; Pearce and Denton 2011; Pearce, Hardie and Foster 2013). Factors such as belief, emotional connection to God, spirituality, religious behavior, or institutional affiliation account for different dimensions of religiosity. People might be highly religious along certain dimensions and less religious along others. Against this backdrop some argue that religiosity can only be satisfactorily captured by typological measurement approaches, since individual combinations of religious dimensions lead to individual profiles of religiosity (see for example Storm 2009, Pearce, Hardie and Foster 2013). The typological approaches of previous research enabled the discovery of some very specific profiles of religiosity. However, comparative research has since shown that these observable dominant religious profiles are not necessarily the same across countries (see for example Davie 1994; 2000; Storm 2009; Voas 2009).

For meaningful comparisons data must be truly comparable, therefore, it is necessary to test data for measurement invariance (see Vandenberg and Lance 2000; Meuleman and Billiet 2012; Davidov et al. 2014; to mention but a few). Invariance tests are particularly important for cross-national data, since respondents have been socialized in different cultural contexts, live in different economic circumstances and speak different languages - all factors that might lead to dissimilar and varied understandings of survey questions and related concepts and thereby compromise data comparability across countries.[28]

However, in the case of religiosity there may be a second element causing structural differences in the data. As religious doctrines play a crucial role in shaping religious beliefs and behavior (Halman and Draulans 2004; McQuillian and Gehrmann 2017), both country-specific differences and individual religious affiliation, i.e. the religion or denomination the individual affiliates with (or the lack of any religious affiliation) can be assumed to affect religious profile development.

[28] Another question in this context, which is not the focus here, is whether survey instruments are translated accurately so as to be understood equally by respondents within and across countries. This kind of measurement error is in principle always possible, however, not too likely due to the high standard of the ISSP translation procedures.

This paper examines cross-national profiles of religiosity from a methodological and substantial angle, beginning with an analysis of whether there is one valid typology worldwide, or if some countries show profiles more similar to one another than to others. Then, it tests the results for invariance in an effort to examine whether it is actually appropriate to compare the empirically found typologies of religiosity cross-nationally. Finally, it looks at the impact of respondents' denomination as a distinguishing element for profile development. The data basis for this endeavor is the ISSP (*International Social Survey Programme*) of 2008, which provides a wide range of variables on religious beliefs, attitudes and behavior for a total of forty-four countries worldwide.

The next section summarizes the main arguments for using typological measurement approaches to measure religiosity and introduces some relevant profiles of religiosity discovered by previous research. This section is followed by theoretical considerations on what kind of influence religions or denomination, respectively, might have for profiles of individual religiosity. Section 4 outlines the central research questions and hypotheses, while section 5 describes the data basis, variables and methods used, followed by results and discussion. The final section concludes the paper with a summary of the main results and their implications for further research.

3.2 Profiles of Religiosity

An established procedure for measuring religiosity in and across societies is the utilization of a typological approach examining the nature of individual religiosity which is then cumulated on a society level (see for example Chaves 2010; Davie 1990; 1994; 2000; Edlund 2013; Jones et al. 2011; Pearce and Denton 2011; Stolz et al. 2014; Voas 2009). This approach is based on the idea that religiosity is a multidimensional phenomenon characterized by multiple factors such as belief, emotional connection to God, spirituality, religious behavior, or institutional affiliation. Unlike continuous measures[29] that take only one specific value indicating an individual's degree of religiosity within a range, typological measures grant all factors the right of a parallel existence and thereby indicate individual religious profiles. A score along a continuous religiosity measure cannot reveal which factors are actually crucial for achieving this score. They merely indicate a certain degree of religiosity, but variations in religiosity, so it is argued, are often differences in pattern rather than degree (Storm 2009, 716) and patterns only appear due to typological measures.

[29] Continuous measures can be, for example, factor scores resulting from a Confirmatory Factor Analysis.

Different dominant profiles of religiosity[30] have been identified across societies by various studies: Edlund (2013) uses ISSP data across twenty-nine countries and three different points of time between 1991 and 2008. His analyses reveal a typology of *Traditional Believers, two different types of Alternative Believers and Non-Believers* – a set of profiles that can be found across countries and time. However, due to the article's approach, this is based on a rather limited number of just three items measuring religious belief.

Stolz et al. (2014) also developed a typology on the basis of ISSP data. The national Swiss study identifies typologies on two different levels of detail along the axes of institutional religiosity and alternative spirituality. The broader classification comprises an *institutional type*, a *reserved type*, a *secular type* and an *alternative type*.[31]

By means of national data, Pearce and Denton (2011) identify five common profiles of religiosity among adolescents in the United States; labeling these the "Five A's": *Abiders, Adapters, Assenters, Avoiders and Atheists*. *Abiders* display high levels of religiosity, while *Atheists* score very low on all measures of religiosity. The three other groups are located between these extremes, each showing certain specific profiles of religiosity.

Jones et al. (2011) look at the impact of gender and ethnicity on religious profiles among young adults participating in two waves of an US-American longitudinal survey. Next to the obligatory extreme groups of *generally low* and *generally high* religiosity, on the basis of the included items they differentiate between a *personal experiential* group, a *personal ritual* group, an *involved* group and a *spiritual-not-religious* group.

Voas examines religious change in twenty European countries utilizing ESS (*European Social Survey*) data. His findings are not that different: Next to one quite consequently *religious* and one *non-religious* group, he discovers a large group of people situated between these two extremes. Instead of differentiating specific profiles among those neither holy religious, nor non-religious, however, he speaks of one group showing "Fuzzy Fidelity" (Voas 2009).

Others have also focused on these "fuzzy groups" which exist between the religious and the non-religious. Davie (1990; 1994; 2000) identifies persons classifiable as *Believing without belonging* in Great Britain as well as groups of those *Belonging without believing* predominantly in Scandinavian countries.

Storm (2009) examines data across ten European countries and identifies Davie's *Believers without belonging*, those *Belonging without believing* as well as four "fuzzy" groups.

[30] The procedure to identify profiles of religiosity is explained in detail in the methods section.
[31] Translation from German by the author.

Siegers (2012) examines the nature and distribution of alternative spiritualities in Europe. Using data from the EVS (*European Values Study*) his analyses reveal six classes of religious orientations across twenty-seven European countries: *atheism, religious indifference, moderate religiosity, individualistic religiosity, church religiosity* and *alternative spiritualities*[32].

With the exception of Siegers, none of these studies systematically tested the latent classes nor the items involved for invariance across groups, leading to the question as to whether the meaning of the latent classes is actually comparable across countries. After a section theoretically exploring the question of what sort of impact belonging to certain religions/denominations might play in individual religious profile formation, this paper then follows the typological measurement approach for the identification of cross-national religious profiles and tests for invariance.

3.3 Religion and Denomination

Religions or denominations[33] provide their followers with an ideology and a code of behavior, more or less strictly formulated depending on the nature of denomination and the level of secularization within the country. The following explores religious attitudes and behavior theoretically expected of various religious groups:

Particularly in Northern and Western Europe those who consider themselves as *not belonging to any denomination* are an already large and steadily growing group. Although parts of this group might have been socialized in a religious context and are more or less affected by the influence of societies' dominant denominations, of all respondents they are expected to show the lowest levels of religiosity across all dimensions. An exception could be a group of *Believers without Belonging* as observed by Davie (1990). Such a group would form a cluster with low levels of institutional religiosity, but with quite strong belief in God and other sacred matters.

Catholics are generally presumed to be more religious than *Protestants* since the secularization process has affected the major Protestant Churches more strongly than the Catholic Church (Therborn 1995, 274). Especially in terms of institutionalized, church-related behavior, Protestants are generally presumed to show lower levels of commitment. The roots lie in the theological foundations of denominations themselves. Individualism receives greater emphasis in Protestantism, since emphasis on the role of the Church as mediator between the believer

[32] Translation from German by the author.

[33] Due to the data basis this paper differentiates between followers of main religions, in the cases of Jews and Muslims. In the case of Christians, it further differentiates between the denominations: Roman Catholicism, Protestantism and the Orthodox Church. To save space, from now on the term denomination will be used to refer to these Christian denominations as well as the main religions.

and God is reduced (Halman and Draulans 2004, 286). Protestants, however, less institutionally involved, might stress their sense of spirituality more than Catholics do.

Theologically, *Orthodoxy* also stresses the community idea, however, unlike in Catholicism, not in a centralistic way "from above", but more in an emancipated manner as "the Church as a whole" (Halman and Draulans 2004, 266). The Orthodox Church puts specific emphasis on ceremonial religious rituals at religious holidays, which creates a number of "occasional church goers" in Orthodox societies (Höllinger 2013, 55). Apart from the theological foundations, one cannot ignore the influence that suppression by the Soviet regime had on the Orthodox Church in affected countries, which is most likely the reason for the rather high level of secularization (Halman and Draulans 2004, 266).

Jewish tradition in general is focused much more on religious practice than on religious belief (Cohen, Siegel and Rozin 2002). In a series of studies Cohen et al. (2002) have shown how also Jewish participants' self-rated religiosity is predicted by their extent of practice but not by knowledge of Judaism or religious beliefs. The reason for this may be that the level of attention Jews give to individuals' mental states is lower compared to other religions (Cohen, Siegel and Rozin 2002). While, for example, Protestants' individual beliefs conceptualize a person's religion, Judaism is passed on via the mother, independently from the individual's connection with God. Or, while for a Protestant having thoughts about immoral actions already means having done something wrong, for Jews, such thoughts are morally neutral (Cohen and Rozin 2001).

Formulating expectations for a Muslim typology of religious attitudes and behavior is more complicated, since to date, this remains less widely examined. However, given that for Muslims being faithful is self-evident (Pace 1998, El-Menouar 2014, 56), high levels of religiosity are expected to be observed on all belief questions. "Church" (or rather mosque) attendance could be comparably low, first, because it is mostly just men who attend, and second, even for men mosque attendance is not an inherent part of Muslim piety as such and therefore, not compulsory (El-Menouar 2014, 55).[34]

3.4 Hypotheses

This paper aims to develop a cross-national typology of religious profiles including thirty-five countries across the world that produces results which pass an invariance test. The database

[34] The implications of this issue for the comparability of the measurement instruments are further discussed in section 5.

comprises some very dissimilar countries. Dissimilarity in this context refers to the languages spoken, the denominational structure, the economic and political situation, recent history, as well as the Church-State relationship, just to mention a few factors. Finding dominant profiles of religiosity that will be valid across a large number of these partly very dissimilar countries, therefore, seems challenging at best. Consequently, the central hypothesis is:

H1a) It is not possible to conduct a valid typology for religiosity including thirty-five countries.

Some countries, however, appear more similar to each other than others. The level of similarity probably increases, if the whole set of countries is broken down into smaller groups.

H1b) The smaller the groups of countries and the more similar the countries contained in these groups, the more invariant the data becomes.

Denominations provide their followers with an ideology and a code of behavior. It is therefore assumed that members of certain denominations are clustered by certain religious profiles reflecting the characteristics of these ideologies and behaviors. Based on the theoretical expectations explored in section 3 the following hypotheses on the specific denominations will be tested:

H2) Those, who do not belong to any denomination show a profile with a low general degree of religiosity across all religious dimensions.

H3) Catholics show a profile with generally high degrees of religiosity across all religious dimensions.

H4) Protestants show a profile that emphasizes belief and spirituality more than institutional commitment.

Hypotheses for Orthodoxy are twofold. Based on religious doctrine, a profile uniting Orthodox Christians could be characterized by rather high church attendance in relation to comparably low levels of belief:

H5a) Orthodox Christians show a profile that emphasizes church attendance more than belief and spirituality.

However, the Socialistic history of Orthodox countries might have caused a general decrease in religiosity that includes church attendance. The second hypothesis therefore is:

H5b) Orthodox Christians show a profile of comparably low levels of religiosity across all religious dimensions.

H6) Jews show a profile that emphasizes institutional commitment more than belief and spirituality.

H7) Moslems show a profile with a generally high degree of religiosity, except for church attendance.

3.5 Data, Variable Selection, and Methods

3.5.1 Data Source

The data used for this study are taken from the ISSP 2008 *Religion III* module (ISSP Research Group 2012) and a cumulated data set containing data from four non-ISSP-member countries (ISSP Research Group 2013) where the Religion survey was conducted in the context of the *Religion Around the World Study* of the 2008 International Social Survey Programme. Together these datasets provide national data from forty-four countries across six continents, thirty-five of these are included in the following analyses.[35]

3.5.2 Variables

The item selection orients on the choices made by Pearce, Hardie and Foster (2013) who conducted an LCA using a wide range of items to identify profiles of religiosity in a US-American national sample. Their choice of items was based on a tri-dimensional religiosity classification scheme[36]: cognitive, affective and behavioral (Cornwall et al. 1986 in Pearce, Hardie and Foster 2013, 62); addressing three general components: religious belief, commitment and behavior. This study covers the same dimensions with similar items[37], offering a rich basis of religious aspects upon which to build typologies.

The cognitive dimension of religiosity is reflected through belief or, conversely, doubts in the existence of God and/or other elements of religion (Pearce, Hardie and Foster 2013, 58). Variables included in this study measure cognitive aspects of religiosity by asking about the extent to which respondents believe in God; if they regard God to be a personal God or, more

[35] For the full list of countries and sample sizes see Appendix II. Data from the Netherlands and Russia are excluded due to data inconsistencies. South Africa, Switzerland, Spain and Northern Ireland are excluded, because of filter errors.
The four Asian countries Japan, South Korea, Taiwan and Sri Lanka are excluded for comparability reasons (see further explanation in the section *Variables*). For Belgium only the regional subsample of Flanders is available.

[36] The purpose of this paper is to a large degree methodological; therefore instead of discussing the single dimensions in great detail, this paper takes on existing interpretations and realizes them with similar items. For a thorough discussion of religious dimensions see for example Glock 1962 and Cornwall et al. 1986. For the variable choice of their measurement see Pearce, Hardie and Foster 2013.

[37] For question texts and answer categories see Appendix I.

transcendent as a Higher Power of some kind. Respondents can also express a more agnostic view, stating they do not know whether there is a God, and not believing there is any way to find out. Another item covers the issue of whether respondents believe God to be concerned with humans. The third measure asks about belief in life after death.

The affective dimension of religiosity captures the emotional connection between an individual and sacred or religious matters (Pearce, Hardie and Foster 2013, 58). One question included for this dimension asks whether life is meaningful because God exists. Other items deal with spirituality, asking the respondents whether they perceive themselves as religious and/or being spiritual, interested in the sacred or the supernatural and thereby differentiate between organized church religiosity and a personal connection with the divine. Because of its big explanation potential, the variable is included although there is some indication that the cross-national interpretation of "spirituality" could be difficult (Siegers 2013, 180).

Religious behavior is addressed by the classic question on public religious practice: frequency of church attendance; and one question on private religious practice: frequency of praying.

In addition to the very substantial aspects of religiosity listed above, a central aspect of this paper is the role a person's denomination plays for individual religiosity profiles. The variable covers seven groups: 1. those reporting to have no religion, 2. Roman Catholics, 3. Protestants, 4. Orthodox Christians, 5. other Christian religions, 6. Jews and 7. Muslims. [38]

Almost all variables used for the measurement of religiosity assume the core element of a monotheistic God. For the measurement of religious behavior, some kind of house of worship where sermons (in some form) are offered to the public at least on a weekly basis is essential. Since this is not necessarily the case in Asian religions, such as Buddhism and Hinduism, the countries where these are the (pre-)dominant denominations, Sri Lanka, Japan, Taiwan and South Korea[39] are excluded from the analyses. Similarly, all Buddhists, Hindus and followers of "other Eastern Religions" in the remaining countries, are equally excluded from the analysis. For all countries the design weight is applied, if provided in the data. A population weight for

[38] Due to the denominational composition of the countries, the broad categories do not necessarily contain the very same denominations in every country. E.g., the category "Protestants" might cover predominantly mainline Lutheran Protestants in one country and mainline Calvinists in others, whereas in all countries also the rather extreme Protestant Free Churches are also covered by this category. For the national-specific composition of the category "Protestants", see the national ISSP Variables "nat_Relig" on ZACAT (http://zacat.gesis.org).

[39] South Korea is a border case here. There are more Christians in South Korea than in the other excluded countries. However, of all religious South Koreans in the sample, 42% do not follow a monotheistic religion. 40% of the whole South Korean sample is non-religious. Since a great number of those have been socialized rather in non-Christian than in a Christian context, South Korea does not meet the selection criteria.

Germany ensures the correct relation between East and West.[40]

Of course, it could be questioned how well some of the measures used can possibly work in the remaining countries in terms of their cultural context or dominant religion, e.g., how well the measurement instruments, designed to measure religiosity in Christian societies, translate into the terminology of other religions, such as Islam and Judaism. A primary example for this possible misalignment is the "church attendance" measure when applied to Islamic societies. Since women rarely attend mosque in the majority of Islamic societies, they consequently answer negatively to questions on church attendance. As a result there is the risk they will be perceived as being significantly less religious than they actually are. In the Turkish ISSP survey, the only homogeneous Islamic society included in the ISSP, instead of asking about Mosque attendance, the *Salah* (5x daily prayer ritual) was queried, which can be performed by both men and women. Whether this concept is truly functionally equivalent and whether the answer scales are then comparable with "church attendance" is debatable. Nevertheless, the item manages to capture religious behavior for women, even if there is some overlap with the question on the frequency of prayer. Despite, or even because of the reasonable suspicions that Christian measurement instruments cannot easily be applied to Muslim or Jewish religiosity, a decision was made to keep Turkey and Israel and the cases of Moslems and Jews in other societies in the analysis.[41] The reason is that these societies, or cases, are actually often included in comparability studies and conclusions are drawn. The following analyses comprise empirical testing of whether such comparisons are actually valid.

[40] Germany is treated as one unit, although it is well known that East and West Germany vary greatly in religious matters. Nevertheless, the country has been united for 25 years now. It shows cultural facets that might have their origins in different regional history, just as other countries do.

[41] For a discussion on how well Western items work for Muslim Religiosity see El-Menouar, 2014; for a discussion on the predictability of response patterns on Western items in non-Western societies see Bechert and Edlund 2015.

3.5.3 Latent Class Analysis

Latent Class Analysis identifies sets of discrete, mutually exclusive latent classes by recognizing similar individual response patterns based on the items included in the analyses. Therefore, it is the method of choice for identifying profiles of religiosity. The analysis assigns each individual to a specific cluster consisting of members sharing the high possibility for providing answers to the survey questions in a certain manner. In other words, they show similar religious profiles. Varying numbers of dominant religious profiles are possible across countries.

To test *H1a* the initial step of analysis[42] is to conduct an exploratory LCA across all thirty-five countries.[43] The 35-country LCA model contains the seven indicators and the variable *country* as a covariate. Both, the indicator variables and the latent variables are treated as nominal. Ten models are run for the data on the integrated dataset with all countries and with the number of clusters being increased by one in each model. Determining the model, i.e. number of classes that represent the data best is necessary to correctly interpret the set of profiles. The most suitable criterion in this case is BIC (L^2), since it accounts for the large sample sizes.[44]

The LCA model is tested for invariance as follows: a structurally homogeneous model is tested against a partially homogenous model and a heterogeneous model. The heterogeneous model allows country effects for both, the intercepts and the slope parameters. The partially homogenous model allows only for country-specific intercept parameters. Interaction effects between country and latent variables (the latent classes), however, are not allowed, that means the slope parameters are set to zero across groups. Finally, for the structurally homogeneous model intercepts and slope parameters are fixed. If the structurally homogeneous model proves to fit the data best, national profiles are structurally equivalent. If the partial homogeneous model fits the data best, country-profiles can still be regarded as comparable. This is conceptually similar to a metric equivalence model in an MCFA context (Kankaraš, Moors and Vermunt 2011, 369). If, however, a fully heterogeneous model fits the data best, profiles cannot be compared, since all measurement model parameters are group specific.

To make sure bigger national samples do not dominate the LCA results over smaller samples, the sample sizes are all weighted to the same size (see also Kankaraš, Moors and Vermunt

[42] Software used: syntax version of Latent Gold 5.0.

[43] Despite very powerful computing technology (4 CPU, each double core; 8GB RAM) it was not possible to accomplish such a huge LCA with the default Latent Gold settings. Following the advice of the Technical LG Guide (2005, 38) the parameters to be calculated were reduced by the standard error and Wald statistics and the calculation was accelerated by only using EM algorithm and not Newton-Raphson. Differences in the results e.g. for the BIC L^2 value due to this procedure manifest just on the second or third decimal.

[44] For a discussion on model selection criteria see for example Bacher and Vermunt 2010.

2011, 374); the median sample size across all thirty-five countries after weighting (N per country after weight=1232)[45].

In the event that the invariance test does not prove structural homogeneity, to test *H1b*, subgroups of countries based on the structure of their national LCA profiles will be built to see if the level of homogeneity can be increased.

3.6 Results

3.6.1 Results I: 35-country LCA: Profile Description and Invariance Test

The BIC L^2 value for the models shown in Table 1a is lowest in the three-class model, but it is not significantly higher in the four-class model. Since any three-class model is equal to a four-class model with zero probability in the fourth class, the decision was made in favor of the four-class model (Table 1c) to give the data a bit more room to unfold.

Table 1a: Statistical parameters for the 35-country LCA homogeneous models

4cluster	**Npar**	**L^2**	**BIC (L^2)**	**df**
Model 1	875	241120.17	-102134.77	32917
Model 2	1785	170437.82	-163327.66	32007
Model 3	2695	146902.20	-177373.82	31097
Model 4	3605	138580.82	-176205.75	30187
Model 5	4515	134173.67	-171123.45	29277
Model 6	5425	130873.16	-164934.50	28367
Model 7	6335	128004.28	-158313.92	27457
Model 8	7245	125764.98	-151063.77	26547
Model 9	8155	123786.35	-143552.95	25637
Model 10	9065	121881.40	-135968.45	24727

Note: For the number of cases included in the single LCAs, see Appendix II.

Table 1b: 35-country LCA invariance test

4 classes	**Npar**	**L^2**	**BIC (L^2)**	**df**
Heterogeneous	205	176632.79	-173608.88	33587
Part-Homogeneous	1055	147029.99	-194347.91	32737
Homogeneous	3605	138580.82	-176205.75	30187

Table 1b shows that compared to a heterogeneous and structurally homogeneous model, the partial homogeneous model fits the data best. This indicates that despite all heterogeneity across countries country profiles can be compared, but structures are not equivalent. Thinking

[45] The calculations have shown that weighting actually does not impact the results very much.

of the different specific religious profiles observed in previous research, the question that arises when looking at Table 1c is: Is there a specific profile that can be recognized?

Table 1c: 35-country LCA – partial homogeneous model, numbers in % (conditional response probabilities*100)

	Non-Religious	Moderately Non-Religious	Moderately Religious	Very Religious
Cluster size	16	19	31	35
Belief in God				
Strong belief, no doubts	0	5	54	92
	1	17	30	5
	2	22	9	1
	11	37	7	1
	24	16	1	0
No belief	63	3	0	1
Life after death				
Yes	3	10	30	71
	6	29	43	17
	23	38	15	4
No	68	23	12	9
God concerns himself with humans				
Agree	0	2	17	65
	0	9	51	23
	3	37	20	3
	26	37	10	5
Disagree	71	15	2	4
Life is meaningful because God exists				
Agree	0	0	6	44
	0	2	27	32
	1	18	31	11
	16	44	29	9
Disagree	82	36	7	4
Religious and/or spiritual				
Religious and spiritual	0	2	21	62
Religious	2	31	60	28
Spiritual	12	38	16	8
None	85	29	3	2
Church Attendance				
Often	0	0	18	59
	0	2	21	20
	20	57	48	17
Never	80	40	13	4
Praying				
Often	0	3	26	56
	0	9	40	38
	5	48	30	4
Never	95	40	5	1

Note: For the color legend for Tables 1c and 3 see Appendix III.

The answer is: no. Instead of showing specific profiles the classes can be ordered – ranging from low levels of religiosity up to high levels of religiosity; from a profile that can be described as *Non-Religious*, to a *Moderately Non-Religious* and a *Moderately Religious*, on up to a profile of *Very Religious* respondents. With every profile the response probabilities increase towards

higher levels of religiosity almost evenly across items. That means the profiles can be organized on a low-high continuum according to the general strength of religiosity across items. The *Very Religious* profile is with 35% the largest, followed by 31% of those being *Moderately-Religious*, 19% *Moderately Non-Religious* and 16% *Non-Religious*. Ordering the profiles along the low-high continuum makes it possible to see diagonals of response probabilities (see Table 1c), which visualize the legitimacy of this ranking. For example in the first item battery dealing with belief in God, in the *Non-Religious* profile the vast majority of 63% does not believe in God. Across the *Moderately Non-Religious* and the *Moderately Religious* profile the majority then shifts from not believing in a personal God, but in a Higher Power, towards believing, but having doubts, up to finally 92% of those in the *Very Religious* profile knowing without any doubts that God really exists. Similar distributions can be observed for the other items.

To investigate whether structural homogeneity can be observed at all across countries, single LCAs with the same items[46] included are conducted on the basis of the national data sets. A careful analysis of the national LCA patterns identifies the countries with the strongest similarity.

3.6.2 Results II: Finding a Structurally Homogeneous Subgroup

The LCAs on the thirty-five national datasets reveal patterns of two to six national profiles of religiosity.[47] Some of these profiles appear more similar than others, as expected. During the investigation-process the patterns were explored very carefully for similarity. In the end the numbers of classes and the structural similarity of response probabilities' patterns were used as criteria for similarity. For this similarity of patterns, it is the characteristics of a cluster that matter, not its size. Let us assume in two different countries three dominant religious profiles appear and in all profiles the probabilities that respondents answer to the survey questions in a certain manner appear quite similar. Then, these two countries are attested similarity, independently of the size of the profile. Following the suggestion of Rudnev, Magun and Schmidt (2014, 11) as the statistical criteria for the similarity of patterns the correlation of profiles is used to confirm or contradict the similarity assumption. Profile by profile for every country, the response probabilities on the items are correlated with the respective response probabilities of items in all other countries that seem to show similar patterns. To count as

[46] For later analyses in the national LCAs also the respondents' denomination is included as an inactive co-variate.
[47] All LCA results are available from the author on request.

similar, only positive correlations above .5 are accepted. From the single profile correlation coefficients the average correlation for each profile across all countries in the group is calculated. Then, once more, the average of these class specific correlation coefficients is ascertained to offer a quality criterion for the similarity of a whole set of response patterns. In a nutshell, the assignment is based on two criteria: 1) the BIC L^2 value for identifying the appropriate number of classes (religious profiles), and 2) systematic correlation tests of response probabilities. These are the countries with the most striking similarity of religious profiles:

Belgium (Flanders), Denmark, Finland, Germany, Hungary, Latvia, New Zealand, Norway, Slovenia, Sweden, Ukraine

For all these countries a four-class solution fits the data best. The four religious profiles show a strong resemblance to the four profiles identified by the 35-country LCA. In all countries a *Non-Religious*, a *Moderately Non-Religious*, a *Moderately Religious* and a *Very Religious* class can be identified.

Table 2: Average Pearson correlation coefficients for each profile class across countries

	Non-Religious	Moderately Non-religious	Moderately Religious	Very Religious
Ø .85	.95	.81	.80	.84

Table 2 reports the correlation coefficients for each profile and the compact average correlation coefficient. Item response probabilities correlate in the *Non-Religious* profile with the exceptionally high coefficient of .95. For the other three profiles they still correlate .80 or higher, which results in a pretty robust average correlation coefficient of .85 for the profiles across the subgroup of countries.

The eleven countries of the identified group are mainly northern and central European. The outer boundaries are Hungary and Slovenia in the south and Latvia and the Ukraine in the east. The only real geographical outlier is New Zealand. What could be the crucial macro indicator responsible for showing similar profiles across certain countries? At first glance, countries appear quite heterogeneous in terms of their denominational structures. In the Nordic countries Protestantism prevails, however, it is commonly understood that the majority we are talking about here are rather secular Protestants (Niemelä 2015; Halman and Draulans 2006), who regard church affiliation more as an "expression of solidarity with society and its basic values" (Hamberg 2003, 50). In Belgium, Hungary and Slovenia the majority of people are Catholic, and in the Ukraine predominantly Christian Orthodox. In Germany, Latvia and New

Zealand, however, those who indicate not belonging to any denomination form the biggest group. All together 25% of respondents across this group of countries are non-religious. So, is the uniting element the high level of secularization? At least not exclusively since e.g. the Czech Republic, the least religious country of the ISSP sample shows entirely different profiles and is therefore not part of the group.

Also, there does not seem to be another uniting element, such as cultural closeness or similar recent history, as the exclusively essential factor for the classification. Despite the regional closeness, there are countries having very different cultural backgrounds, recent history, political situations, and experiencing very different levels of affluence.

The four profiles do not reveal specific patterns of particularly high degrees of religiosity with respect to certain dimensions, and particularly low levels on others, but show evenly distributed strength of religiosity across items. Just as the 35-country model and the country-specific models, the classes emerging from the data show a quite distinct ordinal structure and therefore can be ordered on a continuum from low to high levels of religiosity, ranging from a *Non-Religious* profile, to a *Moderately Non-Religious* and a *Moderately Religious*, on up to a profile of *Very Religious* respondents. The most striking difference between the profiles is the general level of religiosity. Again, ordering the profiles along the continuum makes it possible to see diagonals of response probabilities (see Table 3).

Table 3: Subgroup LCA patterns, numbers in % (conditional response probabilities*100)

	Non-Religious	Moderately Non-Religious	Moderately Religious	Very Religious
Cluster Size	26	31	29	15
Belief in God				
Strong belief, no doubts	0	3	33	87
	1	12	42	10
	1	21	11	0
	9	41	13	1
	23	19	1	0
No belief	66	4	0	1
Life after death				
Yes	2	8	17	67
	6	27	47	23
	23	41	23	3
No	69	24	13	7
God concerns himself with humans				
Agree	0	1	8	66
	0	6	45	30
	2	32	31	1
	21	41	14	2
Disagree	77	20	2	1
Life is meaningful because God exists				
Agree	0	0	1	32
	0	1	15	35
	0	10	31	17
	10	40	41	14
Disagree	89	48	12	3
Religious and/or spiritual				
Religious and spiritual	0	2	20	69
Religious	3	25	58	23
Spiritual	11	39	20	7
None	87	34	2	1
Church Attendance				
Often	0	0	6	44
	0	1	11	20
	26	57	65	28
Never	74	42	18	8
Praying				
Often	0	2	15	62
	0	6	36	33
	5	45	43	4
Never	95	48	6	1

For the sake of completeness, it should be asked: Are there any specific profiles emerging from the data we recognize from previous research? As could be expected based on the patterns of the national religious profiles, again, the answer is: no!

Table 4: Invariance test for the subgroup of 11 countries

11 countries/4 cl.	Npar	L^2	BIC (L^2)	df
Heterogeneous	3623	133200.94	-181397.93	30169
Part-Homogeneous	1073	137436.68	-203753.52	32719
Homogeneous	223	167375.31	-182678.66	33569

Here the invariance test reveals no higher level of homogeneity than for the full 35-country model. This means that even when building a subgroup of countries with proven very similar response probability patterns it is not possible to reach structural homogeneity across countries.

To take this experiment even further, the most similar countries were tested pairwise. It turned out that structural homogeneity could be established only in Sweden and Norway (correlation .94). In other very similar countries with only marginal lower profile correlation coefficients only partial homogeneity could be found.

Table 5: Invariance test for the subgroup further reduced: Sweden and Norway

SE/NO/4 cl.	Npar	L^2	BIC (L^2)	df
Heterogeneous	206	5993.07	-4030,51	1362
Part-Homogeneous	131	6130.31	-4445,09	1437
Homogeneous	106	6260.50	-4498.86	1462

3.6.3 Results III: Distribution of Denominations across Profiles

Since we did not find any profiles with specific characteristics, emphasizing certain beliefs or religious behavior, it is also not very likely we would find profiles exclusively representing certain denominations. The profiles indicating degrees of religiosity just enable us to see whether there is a distinct pattern suggesting that members of some denominations are generally more religious than members of other denominations.

Unsurprisingly, the probability that a respondent who declares not belonging to any denomination shows a *Non-Religious* profile is with 61% very high. 29% show a *Moderately Non-Religious* profile, while despite not being a member of any denomination, 7% appear to be *Moderately Religious* and 3% even show a *Very Religious* profile.

Based on the general conclusions this analysis allows, Catholics across countries appear generally more religious than Protestants. They have an approximately 10% lower probability of showing a *Non-Religious* and *Moderately Non-Religious* profile. Conversely, Catholics are with 14% more strongly represented by the *Moderately Religious* profile and with 5% more represented by the *Very Religious* profile.

Orthodox Christians can be placed between these two other Christian denominations. The majority expresses *Moderately Religious* or *Moderately Non-Religious* views. The probability for Orthodox respondents to show a *Non-Religious* profile is lower than for Protestants, the

Very Religious profile, however, is also less populated with Orthodox Christians than with Catholics or Protestants.

Jews distribute across profiles most similarly to Protestants with a slightly lower probability for the *Non-Religious* profile and higher probability of expressing *Moderately Religious* views. Muslims are by far the most religious group followed by the, admittedly rather randomly compiled, group of respondents from "other Christian denominations".

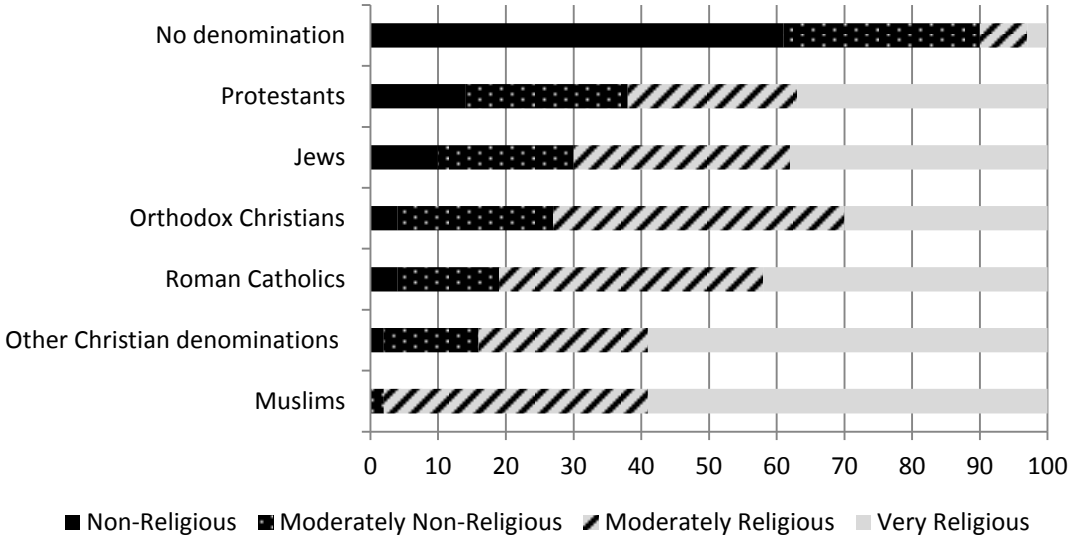

Figure 1: Distribution of denominations in thirty-five countries across religious profiles
Note: Respondents without denomination N=6647, Roman Catholics N=16499, Protestants N=7341, Orthodox Christians N=1938, Jews N=807, Muslims N=3944, Respondents from other Christian denominations N=588.

The data does not produce denomination-specific profiles, but denominations are not evenly distributed across profiles. The assumption that denomination matters seems to be true, though the analysis makes clear that it is not crucial for individual religious profile formation. Turning to the denomination specific hypotheses, *H2* can be confirmed: Those who do not belong to any denomination predominantly show a profile with low levels of religiosity across all religious dimensions. The 10% of respondents with no denomination in the two religious profiles might be called "believing without belonging". *H3* and *H4* cannot be confirmed, Catholics do not cluster in a profile emphasizing institutional commitment and Protestants were not found in a specific profile emphasizing belief and spirituality. The same is true for *H5a*. Orthodox respondents do not cluster in a profile characterized by particularly high church attendance next to rather moderate levels of belief or spirituality. As for *H5b* the question of a yardstick arises. The distribution of Orthodox respondents across profiles shows predominantly moderate religious views and behavior. Socialistic worldviews and/or general

secularization trends might have (had) an impact on the secularization of the Orthodox that places them in between Catholics and Protestants in terms of their general level of religiosity across a sample of thirty-five countries. As a whole, these results appear rather universal and might not do justice to the hypotheses comprehensively. Therefore, they will be further discussed in the following section.

3.7 Discussion…

3.7.1 …on Comparability:

An LCA across thirty-five culturally very dissimilar countries showed a collective set of religious profiles and at least partial homogeneity. The question that arises, however, is what does the invariance test really prove? It serves to validate that the relationships between the latent variables and the indicators are the same across countries. Still, national-specific LCAs have revealed deviant profiles in a number of countries. Can we nevertheless assume that the items on religiosity, really measure the same across Turkish Muslims, Mexican Catholics, and Kenyan Protestants? The answer must be: Measurement validity and conceptual validity are not the same things. Yet across the sample of thirty-five countries single profile structures are so similar that we have good reason to believe that comparisons are actually meaningful. Moreover, the analyses clearly demonstrate that similarity of religious structures need not depend exclusively on societies' cultural closeness or denominational similarity. Achieving structural homogeneity across countries seems almost impossible with a comprehensive set of items that takes into account the different dimensions of religiosity. Reducing the large group down to eleven countries showing structurally very similar profiles still does not produce test results indicating structural homogeneity. Only across just two in many aspects very similar countries (Norway and Sweden) structural homogeneity could be established.

3.7.2…on the Impact of Denomination:

The theory-based expectation was that denominations shape the structure for individual religious profiles. Members of the same denomination were expected to show similar religious profiles. These expectations were only partly met. First, as discussed above, the typology across thirty-five countries did not reveal any specific profiles reflecting characteristics for denominations based on religious doctrine. Therefore, membership on the profiles which appear cannot be clearly predicted by denomination. Only certain trends could be observed: The most religious respondents are Moslems, followed by Roman Catholics, Orthodox

Christians, Jews and Protestants. The vast majority of those without denomination show no signs of religiosity.

One crucial factor affecting the theory-based expectations has not been discussed so far: the country itself. There are actually good reasons to believe that country-specific factors, such as the general level of secularization, the denominational composition of society and the religion-related conflict potential affect the way religion is lived and communicated within the same denominations across countries. Mono-religious societies being rather the exception, most countries differ in their denominational composition. Religious pluralism, however, triggers different social mechanisms, such as religious competition (Iannaccone 1992), religious conflicts, or diaspora effects. In situations of conflict and diaspora effects, religion might even become a badge for ethnic group identity (Bruce 1996). As a result these mechanisms tend to strengthen religious affiliation for the minority group (Cesari 2013). Just to have a glimpse at two examples: contradicting the general pattern, the data on Ireland shows that the 2% of Protestants in the Irish sample have a 50% probability for showing the most religious profile compared to 39% of the Irish Catholic majority. In Mexico the 6% Protestant minority also with 51%, have a higher probability for the most religious profile compared to 34% of the Catholic majority. These are only two examples that feed the theoretical assumption of there being differences between denominations across countries, or, methodologically spoken: an interaction effect of country and denomination.

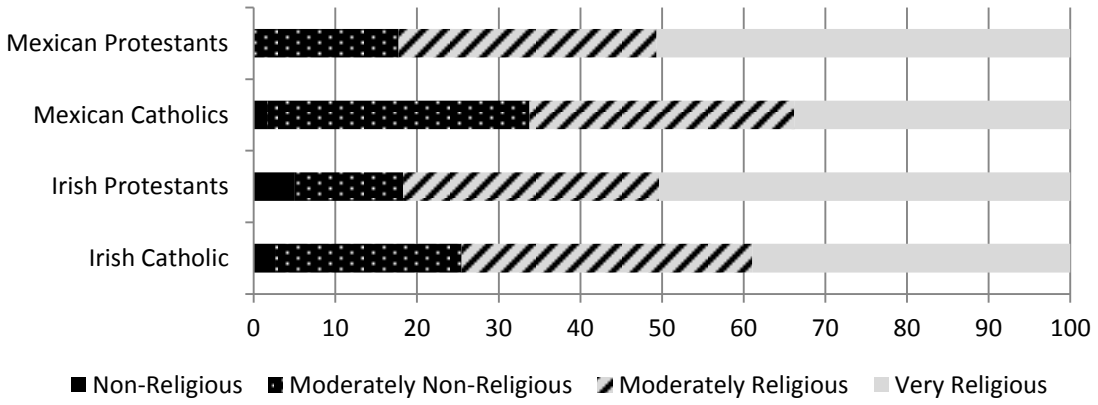

Figure 2: Distribution of Protestants and Roman Catholics in Ireland and Mexico across religious profiles

3.8 Conclusion and Outlook

This paper is based on an examination of data on religious attitudes and behavior collected from thirty-five countries worldwide. Typologies could be empirically derived from this data: one comprehensive typology, and on the basis of a more detailed analysis, a typology on a subgroup of countries showing structurally very similar religious profiles. The analyses have attested to comparability of the data, but no structural homogeneity; and some trends of denominational clustering in profiles were revealed, which is only partly in line with theoretical expectations. Based on these findings future research should make an effort to disentangle possible interaction effects of country-specific macro indicators and denomination influencing individual religious profile development.

One of the most remarkable findings, however, is that the typological measurement approach identifies classes that can be ordered on one latent continuum rather than showing specific characteristics emphasizing single dimensions of religiosity. While Pearce, Hardie and Foster (2013, 57) claim that "humans are rarely consistently low, medium, or high across dimensions of religiosity including institutional involvement, private practice, salience, or belief", the results of this study show that this need not necessarily be the case. Religious profiles can actually be ordered on a low-high continuum according to the general strength of respondents' religiosity. While Storm states "variations in religiosity are often differences in pattern rather than degree" (Storm 2009, 716), the results at hand suggest rather that variations in religiosity are often just that: differences in degree.

What does this finding mean for future research? It suggests that, contradictory to the research opinion presented in this paper and earlier research results, religiosity can actually be measured on one continuum. For the measurement options it means these results provide good reasons to apply ordered latent class analyses (see for example Croon 1990). But also factor analyses, producing continuous factors could be utilized to validly measure religiosity. The consequential next step for research, therefore, is to test multi- and single-item continuous measures of religiosity against the latent classes and compare the explanatory potential of the two approaches.

3.9 References

3.9.1 Data

ISSP Research Group. 2012. International Social Survey Programme: Religion III - ISSP 2008. GESIS Data Archive, Cologne. ZA4950 Data file Version 2.2.0,doi:10.4232/1.11334.

ISSP Research Group; Saflianto, Muhammad; Omondi, Paul; Thavaraja, Joseph; Wanyama, Evangeline. 2013. Religion Around the World Study of the 2008 International Social Survey Programme (ISSP). GESIS Data Archive, Cologne. ZA5690 Data file Version 1.0.1, doi:10.4232/1.11762.

3.9.2 Literature

Bacher, Johann and Jeroen K. Vermunt. 2010. "Analyse latenter Klassen." Pp. 553-574 in: *Handbuch der sozialwissenschaftlichen Datenanalyse.* Edited by C. Wolf and H. Best. Wiesbaden: VS Verlag für Sozialwissenschaften.

Bechert, Insa, and Jonas Edlund. 2015. "Observing unexpected patterns in cross-national research: blame data, theory, or both? Attitudes towards redistributive taxation in thirty-three countries." *International Journal of Sociology* 45(4):327-347.

Bruce, Steve. 1996. *Religion in the Modern World: From Cathedrals to Cults.* Oxford: Oxford University Press.

Cesari, Jocelyne. 2013. *Religion and Diasporas: Challenges of the Emigration Countries.* INTERACT Research Report 2013/01.

Chaves, Mark. 2010. "Rain Dances in the Dry Season: Overcoming the Religious Congruence Fallacy." *Journal for the Scientific Study of Religion* 49:1-14.

Cohen, Adam. B. and Paul Rozin. 2001. "Religion and the morality of mentality." *Journal of Personality and Social Psychology*, 81, 697–710.

Cohen, Adam B., Joel I. Siegel and Paul Rozin. 2002. "Faith versus practice: Different bases for religiosity judgements by Jews and Protestants." *European Journal of Social Psychology* 33 (2): 287-295.

Cornwall, Marie, Stan L. Albrecht, Perry H. Cunningham, and Brian L. Pitcher. 1986. "The dimensions of religiosity: A conceptual model and an empirical test." *Review of Religious Research* 27(3):226-44.

Croon, Marcel A .1990. “Latent class analysis with ordered latent classes.” *British Journal of Mathematical and Statistical Psychology*, 43:171-192.

Davidov, Eldad, et al. 2014. “Measurement equivalence in cross-national research.” *Annual Review of Sociology* 40:55-75.

Davie, Grace. 1990. “Believing without Belonging: Is This the Future of Religion in Britain?” *Social Compass* 37 (4):455-469.

Davie, Grace. 1994. *Religion in Britain since 1945: Believing without belonging.* Cambridge, MA: Blackwell.

Davie, Grace. 2000. *Religion in Modern Europe: A memory mutates (European societies).* New York: Oxford University Press.

Edlund, Jonas. 2013. “Understanding Variation in Religious Believes 1991-2008. The Impact of Modernization and Social Inequality.” pp. 29-46 in *ISSP Data Report: Religious Attitudes and Religious Change*. GESIS Schriftenreihe 13, edited by Insa Bechert and Markus Quandt. Köln: GESIS.

El-Menouar,Yasemin. 2014. “The Five Dimensions of Muslim Religiosity. Results of an Empirical Study.” *Methods, Data, Analyses* 8(1):53-78.

Glock, Charles Y. 1962. “On the study of religious commitment.” *Religious Education. The official journal of the Religious Education Association* 57(4):98-110.

Halman, Loek, and Veerle Draulans. 2004. “Religious Beliefs and Practices in contemporary Europe.” pp. 283-316 in *European Values at the Turn of the Millennium*, edited by Wilhelmus A. Arts and Loe. Halman. Leiden: Brill.

Halman, Loek, and Veerle Draulans. 2006. “How secular is Europe?” *The British Journal of Sociology* 57(2):263-288.

Hamberg, Eva M. 2003. “Christendom in decline: The Swedish case.” in *The decline of Christendom in Western Europe, 1750-2000*, edited by Hugh McLeod and Werner Ustorf, pp. 47-62. Cambridge University Press.

Höllinger, Franz. 2013. “Trends in Church Attendance Among Christian Societies in the Second Half of the 20th Century.” pp. 47-58 in *ISSP Data Report: Religious Attitudes and Religious Change*. GESIS Schriftenreihe 13, edited by Insa Bechert and Markus Quandt. Köln: GESIS.

Iannaccone, Laurence. 1992. “Religious Markets and the Economics of Religion”, *Social Compass* 39: 123–31.

Jones, Janine, et al. 2011. “Ethnic and gender variation in religious involvement: Patterns of expression in young adulthood.” *Review of Religious Research* 53(2):207–25.

Kankaraš Miloš, Guy B. D. Moors, and Jeroen K. Vermunt. 2011. “Testing for Measurement Invariance With Latent Class Analysis.” pp. 359-384 in *Cross-Cultural Analysis. Methods and Applications*, edited by E. Davidov, P. Schmidt and J. Billiet. London: Routledge.

McQuillan, Kevin and Rolf Gehrmann. 2017. “The impact of Religious Denomination on Mentality and Behavior: an introduction.” *Historical Social Research,* 42 (2017) 2:7-22.

Meuleman, Bart, and Jaak Billiet. 2012. “Measuring Attitudes toward Immigration in Europe: The Cross-Cultural Validity of the ESS Immigration Scales.” *Ask: Research and Methods* 21(1):5–29.

Niemelä, Kati. 2015. “No longer believing in belonging – A longitudinal study of Generation Y from confirmation experience to church leaving.” *Social Compass* 62(2):172–186.

Pace, Enzo. 1998. “The Helmet and the Turban. Secularization in Islam.” pp. 165-175 in *Secularization and Social Integration*, edited by Rudy Laermans, Bryan Wilson and Jaak Billet. Leuven: Leuven University Press.

Pearce, Lisa D. and Melinda L. Denton. 2011*. A Faith of Their Own: Stability and Change in the Religiosity of America’s Adolescents.* New York: Oxford University Press.

Pearce, Lisa D., Jessica H. Hardie, and E. Michael Foster. 2013. “A Person-Centered Examination of Adolescent Religiosity Using Latent Class Analysis.” *Journal for the Scientific Study of Religion* 52(1):57-79.

Rudnev, Maksim, Vladimir Magun, and Peter Schmidt. 2014. “The Stability of the Value Typology of Europeans: Testing Invariance with Confirmatory Latent Class Analysis.” National Reasearch University, Higher School of Economics*, Research Paper No. WP BRP 51/SOC/2014.*

Siegers, Pascal. 2012. *Alternative Spiritualitäten: Neue Formen des Glaubens in Europa: Eine empirische Analyse*. Akteure und Strukturen. Studien zur vergleichenden empirischen Sozialforschung 1. Frankfurt/New York: Campus.

Stolz, Jörg et al. 2014. *Religion und Spiritualität in der Ich-Gesellschaft. Vier Gestalten des (Un)Glaubens.* Theologischer Verlag Zürich.

Storm, Ingrid. 2009. “Halfway to Heaven: Four Types of Fuzzy Fidelity in Europe.” *Journal for the Scientific Study of Religion* 48(4):702–71.

Therborn, Göran. 1995. *European Modernity and Beyond: The Trajectory of European Societies, 1945-2000*. London, Sage.

Vandenberg, Robert J., and Charles E. Lance. 2000. "A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research." *Organizational Research Methods* 3(1):4–69.

Vermunt Jeroen K. and Jay Magidson. 2005. *Technical Guide for Latent GOLD Choice 4.0: Basic and Advanced.* Belmont Massachusetts: Statistical Innovations Inc.

Voas, David. 2009. "The Rise and Fall of Fuzzy Fidelity in Europe." *European Sociological Review* 25(2):155–168.

Appendix I

V33: Q16 Closest to R's belief about God

Please indicate which statement below comes closest to expressing what you believe about God.

1 I don't believe in God
2 I don't know whether there is a God and I don't believe there is any way to find out
3 I don't believe in a personal God, but I do believe in a Higher Power of some kind
4 I find myself believing in God some of the time, but not at others
5 While I have doubts, I feel that I do believe in God
6 I know God really exists and I have no doubts about it
8 Don't know
9 No answer

V35: Q18a Belief in life after death

Do you believe in life after death?

1 Yes, definitely
2 Yes, probably
3 No, probably not
4 No, definitely not
8 Can't choose
9 No answer

V42: Q19a God concerns Himself with human beings

Do you agree or disagree with the following: There is a God who concerns Himself with every human being personally.

1 Strongly agree
2 Agree
3 Neither agree nor disagree
4 Disagree
5 Strongly disagree
8 Can't choose
9 No answer

Note: In Indonesia (ID): 5 Strongly disagree was not selected by any respondent.

V44: Q19c Life meaningful because God exists

Do you agree or disagree with the following: To me, life is meaningful only because God exists.

1 Strongly agree
2 Agree
3 Neither agree nor disagree
4 Disagree
5 Strongly disagree
8 Can't choose
9 No answer

V64: Q32 Best describes R

What best describes you:

1 I follow a religion and consider myself to be a spiritual person interested in the sacred or the supernatural.
2 I follow a religion, but don't consider myself to be a spiritual person interested in the sacred or the supernatural.
3 I don't follow a religion, but consider myself to be a spiritual person interested in the sacred or the supernatural.
4 I don't follow a religion and don't consider myself to be a spiritual person interested in the sacred or the supernatural.
8 Can't choose, can't say
9 No answer

Church attendance: Derived from ATTEND*: Attendance of religious services

Apart from such special occasions as weddings, funeral, etc., how often do you attend religious services?

1 Never
2 Less frequently than once a year to several times a year
3 Once to three times a month
4 Once a week or more
** original ISSP variable ranges on an 8 point scale*

Pray – Derived from V59*: Q27 How often R pray

About how often do you pray?

1 Never
2 Less than once a year to once a month
3 2 times a month to several times a week
4 Once a day to several times a day

Note: DK and TR: Max. possible option: Several times a week
** original variable ranges on a 11 point scale*

Appendix II

Country	N original sample	N cases included in national LCA
Australia	1624	1144
Austria	1012	681
Belgium (Flanders)	1232	892
Chile	1487	1284
Croatia	1185	968
Cyprus	998	752
Czech Republic	1475	1224
Denmark	1913	1310
Dominican Republic	2037	1894
Finland	1103	727
France	2356	1611
Germany	1694	1168
Hungary	999	826
Indonesia	1939	1825
Ireland	2046	1687
Israel	1170	878
Italy	1076	913
Kenya	1476	1374
Latvia	1052	877
Mexico	1448	1076
New Zealand	948	686
Norway	1059	652
Philippines	1112	938
Poland	1262	948
Portugal	994	848
Slovak Republic	1119	897
Slovenia	1052	798
Sweden	1211	796
Turkey	1440	1302
Ukraine	1937	1109
Great Britain	1916	1288
Tanzania	1474	1363
United States	1317	1200
Uruguay	990	910
Venezuela	1062	918

Appendix III

Color legend for Tables 1c and 3:

0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%

4. Is there a Continuous Way to God? *A Cross-National Evaluation of the Indicator "Self-Assessed Level of Religiosity" as a Continuous Single-Item Measure of Religiosity*[48]

Abstract

The international sociological research field of religion shares an understanding of religiosity as a multifaceted and multidimensional phenomenon. International survey data nevertheless usually contain only one or two classical indicators for the measurement of religiosity. Given the complex nature of religiosity, this raises the question of how well a single-item measure can explain a multifaceted, multidimensional phenomenon.

On the basis of ISSP data from 2008, this paper tests the validity of the indicator *self-assessed level of religiosity (SAR)* for the measurement of religiosity in international comparative research, by (1) evaluating of the continuous SAR indicator's predictive potential, (2) comparing the results with the likewise continuous, but multi-item SEM - and the multi-item, but categorical LCA approaches, and (3) analyzing whether the predictive potential of the three measures varies across thirty-five countries.

The results reveal a rather unexpected, generally convincing performance of the single-item measure. However, they also make quite clear that the measures' performance is to a great deal context-dependent.

Keywords religiosity, single-item vs. multi-item, LCA, SEM

[48] under review at *Social Compass*

4.1 Introduction

Conducting international social surveys, such as the ESS, WVS or ISSP is expensive.[49] Institutes in countries with very different levels of affluence usually have to raise the financial resources required for conducting their national surveys without funding from the survey program. Each question has to be translated into native languages, has to be asked by trained interviewers, the response has to be coded and a comprehensive dataset then compiled. All these steps are expensive. To study a research topic, however, it is always desirable to have as many measurement instruments available as possible. Consequently, we are facing a dilemma: social scientists rely on getting the greatest variety of measurement instruments to work with across as many countries as possible; yet, the survey needs to be affordable to be conducted at all. A direct result of such economic considerations often is that international surveys covering a variety of topics usually contain a rather limited number of questions on each topic.

An important topic under examination since the earliest days of social science research is religion. Individual religiosity is regarded as an impact factor reflecting attitudes and behavior with respect to various life issues. Economic considerations, however, often present the biggest challenge for survey programs in selecting the most functional indicators for capturing religiosity. The initial aim of this paper is to evaluate the functionality of one of the classical indicators for the measurement of religiosity. It asks the respondents to self-assess their religiosity along a scale. This "self-assessed level of religiosity" indicator will be abbreviated as SAR hereafter for the purposes of this paper.

Why evaluate SAR? The most popular and probably most frequently used single indicator for religiosity is church attendance, since it is a good indicator for the vitality of religious practice (Casanova, 1994), which correlates highly with indicators of religious belief (Jagodzinski and Dobbelare, 1995) and has a high availability across international surveys (Vezzoni and Biolcati-Rinaldi, 2015). There are, however, at least two good reasons why church attendance alone might actually not be an ideal measure for international comparative research on religiosity. First, in times of increasing religious pluralism also in societies of traditionally monotheistic religions, more and more individuals might not regard a church or an equivalent as a core element of their religiosity. Secondly, forms of individual religiosity, frequently dissociated from classical churches, are on the rise in many countries (see, for instance, Beckford, 2003). These individual forms of religiosity cannot accurately be captured by the frequency of any church attendance. SAR, by contrast, is not impacted by any specific religious theology or

[49] ESS – European Social Survey; WVS – World Values Survey; ISSP – International Social Survey Programme

liturgy and, thus, seems to be a much more promising indicator for international comparative research.

Nevertheless, two potential main points of criticism arise when using the SAR indicator as a measure for religiosity. Since few would disagree that religiosity is a multifaceted phenomenon, a common argument is that one single item cannot validly measure such a diverse issue. To address this argument I will compare the predictive potential of SAR with another continuous, yet multi-item measure: the factor scores provided by Structural Equation Models (SEM) based on a variety of indicators covering three general components of religiosity: religious belief, commitment and behavior.

The second point of criticism is the continuity of the measure. Some researchers argue that variations in religiosity are based on differences in pattern rather than degree of religiosity (see for example Davie, 1990; McGuire, 2008; Storm, 2009; Chaves, 2010; Pearce, Foster and Hardie, 2013). They believe that only specific religious profiles of religiosity are truly capable of capturing individual religiosity. These profiles appear when individuals show high levels of religiosity towards certain aspects of religiosity and low levels towards others. To take that categorical approach into account, I will extend the comparison with latent classes resulting from Latent Class Analyses (LCA), based on the same indicators as the SEM models. LCA identifies profiles of religiosity and constitutes multi-item, categorical measures, unlike the SEM. These analyses will be conducted across 35 countries with different cultural background and denominational structures.

In this paper I will (1) evaluate the predictive potential of the continuous SAR indicator, (2) compare the predictive potential with the likewise continuous, but multi-item SEM and the categorical, typological LCA approach, and (3) analyze whether the predictive potential of the three measures varies across countries.

Following this introduction I will outline the main arguments for single-item vs. multi-item measures and continuous vs. categorical measurement approaches; and point out the strengths and weaknesses of self-assessed indicators. I will then formulate my hypothesis based on these theoretical considerations. The succeeding methods section explains the methodological procedures, followed by the results, a discussion of these results and to close out, concluding remarks.

4.2 Measures

4.2.1 Single-Item vs. Multi-Item Measures

The ongoing debate about the trade-offs of single - vs. multi-item measures (hereafter SI and MI) endures in the social sciences and related research fields, such as psychology and medicine, both notorious for their large scales. Its crucial arguments are quite obvious: The main advantage of SI measures is that they are more pragmatic and economic, since they require less space, time and money for the survey. It can also be assumed that it is more convenient for respondents not to answer numerous questions on the very same concept, creating the feeling of repeating oneself - a feeling that might easily lead to drop outs. There is, however, methodological concern regarding the reliability of SI measures, since if there is only this one item it is not possible to measure its internal consistency reliability. Additionally, SI measures are more likely affected by random measurement error (see for example Schnell, Hill and Esser, 2011: 126f.). They are also more vulnerable to unknown measurement bias, caused by context effects due to question order in the questionnaire; a possible scope of translation errors or inaccuracies of functional equivalence in international comparative surveys. These errors, if they affect only single items, have a good chance to be canceled out by multi-item measures, however, multi-item measures definitely cost more space, time and money.

Previous research indicates that there is no definite answer to the question whether SI measures can be as reliable and valid as MI measures. Simulation results in market research produce evidence that under most conditions typically encountered in practical applications, MI measures outperform SI measures in terms of predictive ability (Diamantopoulos et al, 2012). However, across scientific disciplines studies quite frequently provide examples of SI measures being suitable substitutes for long MI scales in terms of reliability, validity and/or predictive ability:

A very prominent example is the successful measurement of the Big Five constructs of personality, which are traditionally measured using very large MI scales, up to 240 items (e.g. NEO-PI-R by Costa and McCrae, 1992), with single items (see Gosling, Rentfrow and Swann, 2003; and Woods and Hampson, 2005). Hoeppner et al. (2011) tested a single item of self-efficacy against a well-established 20-item measure to predict alcohol use relapse across a sample of young adults in residential substance use treatment. They found the SI not only equally suitable, but exhibited superior predictive validity. In the broader field of sociology, Lutz et al. (2013) developed a single item for the measurement of physical attractiveness. To avoid having a great range of items for that measurement, their approach was to set a valid comparative yardstick. Robins et al. (2001) tested a SI measure for self-esteem and found

nearly identical correlations with a wide range of criterion measures as a well-established MI scale.

There are, however, also studies recommending SI measures albeit with constraints. Postmes, Haslam and Jans (2013) test a SI measure for self-identification and find it to have high reliability, high validity and high utility. Nonetheless, they assume that single measures struggle to achieve acceptable reliability, if the constructs they assess are broad and heterogeneous. Consequently, they see the reason why the single item for the measure of social identification works so well in the construct being quite homogeneous. This assumption is not new. In 2002 Loo compared the utility of SI measures with MI measures on attitudes towards "belief in a just world" and recommend that SI measures can be considered only if the single item reflects a homogenous or unidimensional construct (Loo, 2002: 73).

Also Bergkvist and Rossiter (2007), comparing the predictive validity of SI and MI measures on attitudes towards ads and brands in marketing research, noted that the determination as to whether a SI or MI scale should be used is construct-dependent. They suggest that homogeneous, concrete constructs should be measured by SI scales, whereas heterogeneous constructs of an abstract or complex nature should be broken down into concrete components, each measured with separate items (2007: 183). By replicating Bergkvist and Rossiter's results Diamantopoulos et al. come to the similar conclusion that the relative performance of SI measures is context and construct-specific. They warn to abandon established MI scales even for concrete constructs, since in their opinion an SI measure which may perform as well as the MI measure on one context may not do so in another (Diamantopoulos et al., 2012: 446).

According to these recommendations, measuring religiosity with a single item does not seem very promising, since religiosity is certainly not a homogeneous construct, especially not across countries.

4.2.2 Categorical vs. Continuous Measures

The second pair of contradicting measurement ideas addressed in this paper is the utility of categorical vs. continuous measures. It is widely agreed that religiosity is a multi-dimensional concept (see for example Glock, 1962 or Cornwall et al., 1986) and that the differentiation of these dimensions in empirical research is strongly recommended (Kecskes and Wolf, 1995). A plausible argument is that if religiosity is a multi-dimensional phenomenon, with each dimension having its own continuum from low to high, it cannot be measured on one single

continuum (see for example, Storm, 2009, or Pearce, Foster and Hardie, 2013). According to this argument, it is the combination of factors, such as belief, the personal value of rituals, worship or tradition, personal identification, spirituality and institutions (Storm, 2009: 702) creating the profiles that actually represent individual religiosity. In different studies many religious profiles, shared by substantial groups in society and occurring due to such combinations of factors could be identified. Some have been given quite telling names, such as *believing without belonging* or *belonging without believing* (Davie, 1990, 2000, 2007) referring to individuals, who show strong faith in God, but do not attend church on a regular bases and vice versa. Members of another frequently encountered group distance themselves from the classical churches, but are not secular. These people have been called *post-Christian spirituals* (Houtmann and Aupers, 2007) or *Alternative Spirituals* (Siegers, 2012). The term *Sheliaism* was coined by Bellah et al. (1985), referring to a kind of personal common sense religiosity, named after the respondent Shelia who claims to put her faith in her "own little voice"(Bellah et al., 1985: 221). It seems convincing that such special profiles of religiosity, resulting from specific combinations of religious factors cannot be identified with a continuous factor. Or in other words: Measurement on one continuum assumes a singularity that seems out of place for a phenomenon that combines as many elements as religiosity does.

Taking into account the high correlations of these religious factors, however, another research opinion questions the autonomy of religious dimensions. They argue that factors or sub-dimensions actually sum up to one primary dimension, covered by the umbrella-term *religiosity* and accordingly can be measured on a single continuum from low to high.

The examples for continuous MI scales are numerous: In the field of consumer research Wilkes, Burnett and Howell (1986) developed a continuous one factor scale with the four items *church attendance, importance of religious values, confidence in religious values* and a *5 point SAR indicator.* Shukor and Jamal (2013) needed five items (*I believe in Allah, I carefully avoid shameful acts, I always perform my duty as a Muslim, It is important for me to follow Allah's Commandments conscientiously, Religious beliefs influence all my dealings with everyone*) for an equivalent measure for Muslim religiosity.

Huber (2003) developed the Centrality of Religiosity Scale (CRS) in three different lengths of 15, 10, or 5 items covering the five core dimensions as defined by Glock (1962). According to Huber, these dimensions are the basis for a general personal religious construct system that can be measured on one continuum, because the general intensity of the single dimensions allows an estimation of the frequency and intensity of the activation of this general religious

construct system; and the more often the religious construct system is activated, the higher the level of personal religiosity (Huber, 2003; Huber and Huber, 2012).

Using ISSP data, from 1998 and ESS data from 2002, Wolf (2005) found that the three indicators *church attendance*, *praying* and *SAR* all load on one common factor (religiosity) and that they all explain it equally well. He built an index based on the factor scores and arrived at the conclusion that at least traditional Christian religiosity throughout Europe can be reliably and validly measured on one continuum with a set of these three items (Wolf, 2005: 292).

The shortest version of an index was provided by Jagodzinski and Dobbelere (1995), who sought to measure church religiosity on the basis of EVS data from 1981 and 1990. They built an index utilizing *church attendance* and *membership in a denomination*. To find out how well this index explains church religiosity they correlated it with four other indicators for religiosity. Since correlations are high, they concluded that their *church integration* measure is an excellent indicator of church religiosity – indirectly reflecting the degree of individual religiosity (Jagodzinski and Dobbelere, 1995: 90).

The categorical approach criticizes such continuous indices, because if a scale is constructed from multiple variables there are many ways of achieving the same score (Storm, 2009: 704). The information on which concrete variables the respondent scores high and on which low, or in other words: how individuals uniquely combine varying types and levels of religious practices, beliefs, and salience gets lost, and with it the information on the concrete nature of individual religiosity (McGuire, 2008; Storm, 2009; Chaves, 2010; Pearce, Foster and Hardie, 2013). The question then arising is how important is that information for specific analytical purposes?

4.2.3 Self-Assessment Measures

Self-assessment measures can be interpreted as cross-dimensional global measures, since their nature is to summarize diverse components anchored on different dimensions of typically rather broad topics. Classical topics measured by self-assessment indicators therefore usually contain many individual components, as is the case with individual health status or religion. In the health context a self-assessed health item has proven to work as expected by pulling together or summarizing the components that make up the health status domain (Krause and Jay, 1994: 940). In the religion context it is the SAR indicator that measures all those components of religiosity a respondent connects it with and, thus, covers the whole spectrum of individual religiosity. The other side of the coin is that, without any control measures, it

remains unclear which components respondents actually refer to or how different components are weighted when the item is answered. For the single individual that might not be a problem. Hoeppner et al. (2011: 312) conclude that more personification creates more potent predictors (in their case of addicts' probability of a relapse). The question arises, however, to what extent we can speak of comparability across individuals, let alone countries if there is such a large degree of interpretative freedom. Another disadvantage the indicator might involve, particularly for cross-national research, is that it highly depends on reference groups. Moderately religious people in more religious societies may rate themselves as less religious than moderately religious people in more secular societies (Bechert, 2013). In homogeneous societies it is especially difficult to assess one's own position, since if all people are on the same level of religiosity it is not possible to decide whether this is a high or low level (Jagodzinski and Manabe, 2009). Without a comparative yardstick – the ideal would be a macro-variable indicating the societies' level of religiosity/secularization – this problem cannot be solved.

4.3 Research Questions and Hypotheses

How well will the SAR indicator perform, compared to the multi-item measurement approaches of SEM and LCA? The more variables the measure includes, the more aspects of individual religiosity are covered. Despite the convincing performance of SI measures in previous research, I assume that the MI measures still depict the more comprehensive picture of individual religiosity and therefore hypothesize:

H1: The multi-item measures for religiosity (LCA generated profiles and a SEM generated index score) will achieve a higher predictive research value than the single-item SAR measure.

The SEM approach creates a MI measure that operates on one continuum – from low to high levels of religiosity, just like the SAR indicator – only that SAR is not a MI measure. The categorical approach differs here by depicting religiosity on different continuums. Previous research results discussed above lead to the assumption that generally a great deal of information on individual religiosity gets lost if measured on just one continuum. Therefore, I assume that generally, the categorical approach will produce significantly more potent predictors than the continuous measurement approach. H2 therefore states:

H2: The multi-item, categorical (LCA) measure has a greater predictive research value than the multi-item continuous (SEM) measure.

4.4 Data, Country Choice, Variables,[50] and Methods

4.4.1 Data and Countries

The data used for this study are taken from the ISSP 2008 "Religion III" module and a cumulated data set containing data from four non-ISSP-member countries in which the Religion survey was conducted in the context of the "Religion Around the World Study of the 2008 International Social Survey Programme" (ISSP Research Group, 2012; 2013). Together these datasets provide national data from forty-four countries across six continents, thirty-five of which are included in the following analyses. For the full list of countries and sample sizes see the Appendix.[51]

All countries included in the analyses meet one crucial criterion: the dominant religion includes the core element of a monotheistic God. This criterion was set, because it is assumed for almost all variables used for the measurement of religiosity. Since this is not necessarily the case in Asian religions, such as Buddhism and Hinduism, Japan, South Korea, Taiwan and Sri Lanka as well as all Buddhists, Hindus and followers of "other Eastern religions", in other countries are excluded from the analysis. For all countries the design weight is applied, if provided in the data.

4.4.2 SAR

The core indicator for this study measures the respondents' self-assessed level of religiosity (*Would you describe yourself as…?*) on a seven point scale. For the analyses the variable was recoded to range between 0 *Extremely non-religious* and 6 *Extremely religious*.

4.4.3 Variables for the Multi-Item Measures

Item selection for the MI measures is based on a tri-dimensional religiosity classification scheme: cognitive, affective and behavioral (Cornwall et al., 1986); addressing three general components: religious belief, commitment and behavior. The following seven items offer a rich basis of religious aspects upon which to build the SEM models and the LCA typologies.

---

[50] For the original question texts and the precise scales see http://zacat.gesis.org/webview/index.jsp?object=http://zacat.gesis.org/obj/fStudy/ZA4950. Variables were recoded to all point in the same direction: low values = low levels of religiosity.

[51] Data from the Netherlands and Russia were excluded due to data inconsistencies. South Africa, Switzerland, Spain and Northern Ireland are excluded, because of filter errors. For Belgium only the regional subsample of Flanders is available.

The cognitive dimension of religiosity is reflected through belief or, conversely, doubts in the existence of God and/or other elements of religion (Pearce, Foster and Hardie, 2013: 58). Variables included in this paper measure cognitive aspects of religiosity by asking whether the respondents believe in God; have doubts; or do not know whether God exists; if they regard God to be a personal God or, more transcendent, as a Higher Power of some kind; whether they believe God to be concerned with humans and, finally, if they believe in life after death.

The affective dimension of religiosity captures the emotional connection between an individual and sacred or religious matters (Pearce, Foster and Hardie, 2013: 58). Items included ask respondents whether life is meaningful because God exists and whether they perceive themselves as religious and/or spiritual; interested in the sacred or the supernatural. This item thereby differentiates between organized church religiosity and a personal connection with the divine.

Religious behavior is addressed by the classic questions on public and private religious practice: frequency of church attendance; and frequency of praying.

4.4.4 Dependent Variables for the Regressions

The four dependent variables consist of two items that only indirectly relate to religion and two items that directly address religion.

Attitudes towards abortion (item A) and homosexuality (item B) are indirectly connected since Christian churches have a vigorously declared opinion regarding both of these issues. The items ask 1) whether the respondents personally think it is wrong for a woman to have an abortion, if there was a strong chance of serious defect in the baby and 2) whether or not it is wrong for two adults of the same sex to engage in sexual relations.

Items directly connected with religion ask the respondents to indicate their levels of agreement towards the statement “We trust too much in science and not enough in religious faith” (item C) and (item D) asking about their level of confidence in churches and religious organizations.

4.4.5 The Multi-Item Measures

The MI measures are the factor scores of a Structural Equation Model (SEM) run with STATA 14 and the latent classes (religious profiles) generated by Latent Class Analysis (LCA) run with Latent Gold 5.1. Both measures are based on the seven indicators for religiosity explained above.

SEM identifies response patterns on the basis of the items included and assigns them to certain latent factors. Theoretically, various numbers of factors are possible across countries. Exploratory SEMs have shown that in the vast majority of countries the items load on one single factor that could be called *religiosity*. There are only very few countries for which two factors can be observed. I decided to constrain all SEMs to one main factor. This seems to be the appropriate procedure, since this paper, at least marginally, deals with the question of practicability of measures for religiosity. If we want to extract one continuous measure from SEM, the main factor is what we need to work with analog to the SAR measure.

Latent Class Analysis also identifies response patterns on the basis of the items included in this study and assigns each individual to a certain latent class. The members of each class share the high possibility for answering to the items in a certain manner. Each latent class thereby represents a religious profile. Theoretically, various numbers of substantial dominant religious profiles are possible across countries.

In the analysis for each country data set ten LCA models are run separately with the number of clusters being increased by one in each model. Both, the indicator variables and the latent variables are treated as nominal. There are a number of criteria that could be taken into account to decide for the appropriate number of classes, indicating the model that most aptly, represents the data. Considering the large number of countries, however, it is necessary to concentrate on one strong, standardized criterion to distinguish between models. The most suitable criterion here is BIC (L^2), since it accounts for the large sample sizes. Therefore, the models with the lowest BIC (L^2) values are chosen as the best representatives of the data.[52] Cases with any missing value on the seven indicators are excluded from the analysis.[53]

The first analytical step is to now find out how well the three measures correlate with one another, in other words, how much common variance they explain. If they do not correlate there is no justification to use them alternatively for the same measurement purpose. For the two continuous indicators of SEM and SAR R^2 coefficients will give us this information. The

[52] For a discussion on model selection criteria see, for example, Bacher and Vermunt, 2010.

[53] The highest missing rates can be observed on the "Believe in life after death" question and the question concerning spirituality (on average for both variables below 10%). There are no extreme outliers in any country on specific variables. In the Ukraine missing rates are generally rather high.

question how the typological approach relates to the continuous measures will be answered by Variance Analyses. Eta $(\eta)^2$ indicates the effect size, or in other words, the amount of explained variance of the profiles (by the LCA cluster membership variable) on the SEM scores and on the SAR indicator respectively. Additionally, the post-hoc *Scheffe* test will indicate whether the single profiles differ significantly with respect to the mean values of the continuous measures of SAR and SEM. If they do, each profile can significantly be assigned to a mean value of a continuous indicator and the added value of the categorical approach is questionable. If they do not differ significantly, we must assume that there is something else beyond a mean value of general religiosity leading to different profiles that can only be explained by means of categories and not with the continuous SAR and SEM indicators.

The second step on the way to evaluate the functionality of the SAR indicator is to analyze its predictive value for external criteria compared to the two other measures. For this purpose regression analyses will be run with all three measures examining their explanatory potential on attitudes towards abortion, homosexuality, the balance of trust in science vs. faith and confidence in religious organizations. For the LCA measure dummy variables will be constructed from class-membership. Membership to all latent classes, but the most non-religious which serves as the reference category, are included in the regressions as independent variables. All models include the control variables sex and age. To get the net explained values the models are additionally run with only these control variables and the resulting adjusted R^2 is subtracted from the full model's R^2.

4.5 Results

Rather than make this paper excessively long by presenting LCA results with seven indicators from 35 countries, I have chosen to store the single LCA outputs in data depository[54] with free access for those interested in the details.

[54] During the peer review process these results are available from the author on request.

4.5.1 Step One: Correlation, Variance Analyses, and Scheffe Test Results

Table 1: Shared explained variance of the three measures and significance of cluster differences by SAR and SEM factor score means

	1	2 and 3		4 and 5	
Country	R^2 SAR/SEM	Effect size: η^2		✓= All clusters highly significantly different Clu/Clu = these religious profiles (clusters) are not highly significantly different	
		SEM model	SAR model	by SEM	by SAR
Austria	.62	.87	.57	✓	Clu2/Clu3[I]
Australia	.68	.94	.66	Clu5/Clu2[II]	Clu5/Clu2[III]
BE-Flanders	.66	.88	.64	✓	✓
Chile	.33	.73	.28	✓	✓
Cyprus	.55	.87	.46	✓	✓
Czech Republic	.68	.94	.67	✓	Clu3/Clu4[IV]
Germany	.77	.91	.74	✓	✓
Denmark	.63	.92	.61	✓	✓
Dominican Republic	.15	.65	.10	✓	Clu1/Clu3[V]
Finland	.68	.92	.64	✓	✓
France	.68	.95	.68	✓	✓
Great Britain	.69	.93	.70	✓	✓
Croatia	.62	.90	.61	✓	✓
Hungary	.72	.91	.66	✓	✓
Indonesia	.08	.61	.02	✓	✓
Ireland	.52	.84	.49	✓	✓
Israel	.63	.90	.60	✓	✓
Italy	.70	.90	.66	✓	✓
Kenya	.24	.51	.16	✓	✓
Latvia	.67	.92	.60	✓	✓
Mexico	.38	.78	.32	✓	✓
Norway	.62	.92	.58	✓	✓
New Zealand	.65	.94	.62	✓	✓
Philippines	.03	.39	.01	✓	Clu1/Clu2[VI]
Poland	.56	.86	.52	✓	✓
Portugal	.54	.83	.46	✓	✓
Sweden	.64	.92	.58	✓	✓
Slovenia	.71	.91	.67	✓	✓
Slovakia	.77	.95	.74	✓	✓
Turkey	.30	.69	.27	✓	✓
Tanzania	.19	.70	.15	✓	✓
Ukraine	.56	.89	.53	✓	✓
USA	.50	.90	.46	✓	✓
Uruguay	.50	.86	.41	✓	✓
Venezuela	.49	.55	.15	Clu1/Clu3[VII]	Clu1/Clu3[VIII]

Note: All models are highly significant (p=.000). (Exception: The Philippines, SEM: .008)
significance level: [I].655; [II].729; [III].604; [IV].058; [V].124; [VI].008; [VII].115; [VIII].362

The first column (1) in Table 1 shows the R^2 values for the relation of the two continuous indicators SEM and SAR. In the vast majority of countries the explained variance ranges

between 50% and 70%. The highest correlations can be observed in Germany and Slovakia, while in the Philippines and Indonesia there is no real correlation to speak of at all between the measures.

The results presented in the other columns describe the relation of the categorical LCA profiles with the continuous indicators. They are generated by Variance Analyses conducted for each country. F-Tests (not presented in the table) show that the models for all countries, with the exception of the Philippines, are highly significant. That means in each country the latent classes (religious profiles) explain the SAR indicator as well as the SEM factor significantly. These results indicate that the different methods measure the same thing and prove that the methods work reliably across countries.

Eta square (η^2) is an indicator for the effect size. It shows how much variation is explained by the factor in the variance analysis (here the latent classes). Thus, for a single-factor variance analysis η^2 is equivalent with R^2. Column 2 and 3 show that η^2 values are generally high for the models in almost all countries. The effects are, however, higher for the latent classes explaining the SEM factor than the SAR indicator, which is not particularly surprising, since profiles and the SEM factor contain the very same indicators. In 18 out of 35 countries the explained variance for the SEM factors is even 90% and higher. The lowest effects appear in The Philippines, Venezuela and Kenya. Across countries η^2 for profiles/SAR indicator is only very slightly lower than R^2 of SEM and SAR. It still ranges between 50% and 70% in the vast majority of countries. Again, for the Philippines and Indonesia there is almost no noticeable effect.

The post-hoc test now clarifies if there are significant differences between the latent classes with respect to the mean values of SEM and SAR indicators. Columns 4 and 5 indicate whether all latent classes differ significantly with respect to the mean value of SEM (column 4) and SAR (column 5). In the vast majority of countries (29/35) all latent classes differ significantly with respect to both indicators. Austria, the Czech Republic and the Philippines are exceptions here, showing non-significant differences between two classes with respect to the SAR mean value (column 5). These effects, concerning the very same latent classes, are also non-significant in the case of the SEM factor (column 4) only for Venezuela and Australia. The results for the latter are particularly eye-catching, since the general effect size is large for both indicators and the level of non-significance very evident for the two religious profiles (Clu 2 and 5).

4.5.2 Step Two: Regression Results

Table 2 shows the average adjusted R^2 for the regressions of the three measures of religiosity with the four external factors: attitudes towards abortion, homosexuality, trust in science vs. faith and confidence in churches, and religious organization across 35 countries. For all countries the expected direction can be observed: the more religious the individual the less liberal the attitudes towards abortion and homosexuality, the stronger the agreement towards the statement that we should trust more in faith than in science and the stronger the confidence in churches and religious organizations. Two things become very clear, first, religiosity explains more of the directly religion-related items (C and D) than of the rather indirectly related items (A and B). Second, H1 can be confirmed as on the 35-country average the MI measures explain more of the dependent variables than the SI measure. However, the difference is not as pronounced as expected. These results are not too surprising, but nevertheless important to be confirmed.

Regarding H2, on the basis of the summed up average R^2 values we can hardly make a statement on which of the MI measures work better, since the difference is small (1% more explained variance for the LCA profiles on abortion (A), but 2% more explained variance for the SEM scores on the confidence item D). For the other two items the explained variance is exactly the same.

Table 2: Regression Results: Ø adjusted R^2 across 35 countries

	A: Abortion	**B: Homosexuality**	**C: Science vs. Faith**	**D: Confidence in religious org.**
SAR	.05	.05	.11	.18
SEM	.09	.07	.14	.22
Profiles	.10	.07	.14	.20

Figures 1-4 show these regression results separately for each external dependent factor by country. The countries are ordered by the average size of R^2 across all three regressions with the same dependent variable:

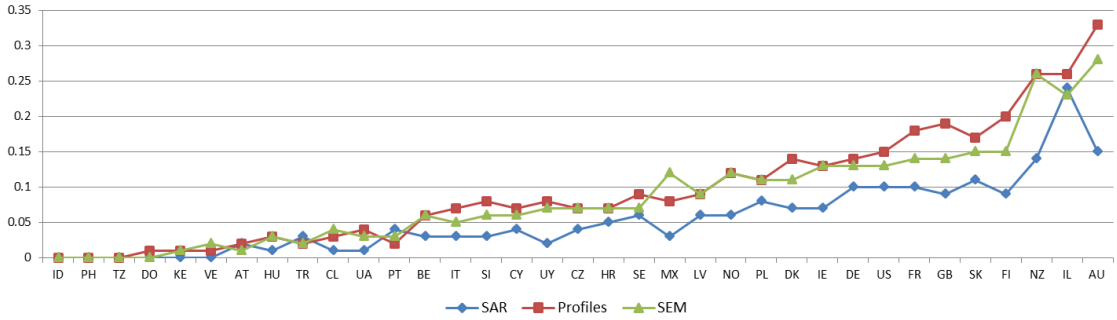

Figure 1: Adj. R^2 values for the regressions SAR/ Profiles/SEM – abortion (A)

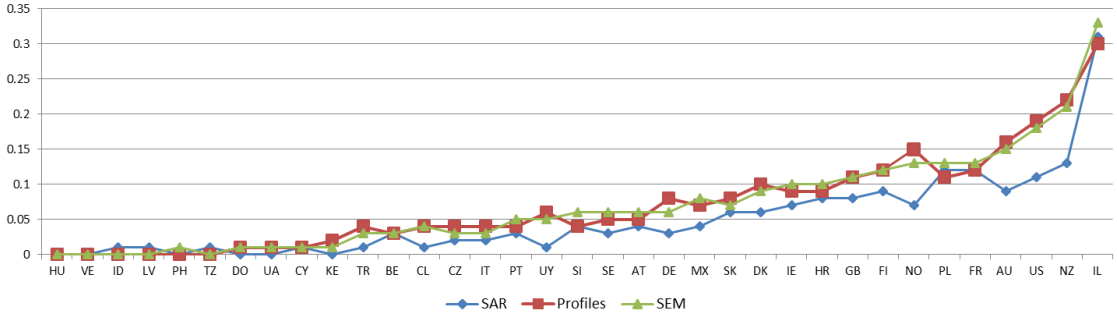

Figure 2: Adj. R^2 values for the regressions SAR/ Profiles/SEM – homosexuality (B)

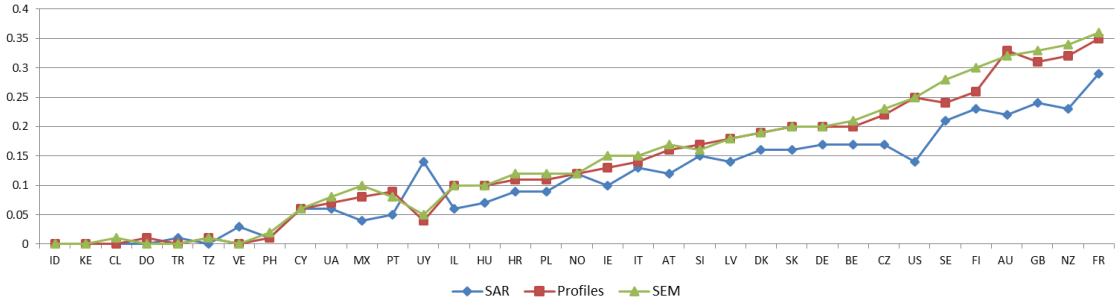

Figure 3: Adj. R^2 values for the regressions SAR/Profiles/SEM – science vs. faith (C)

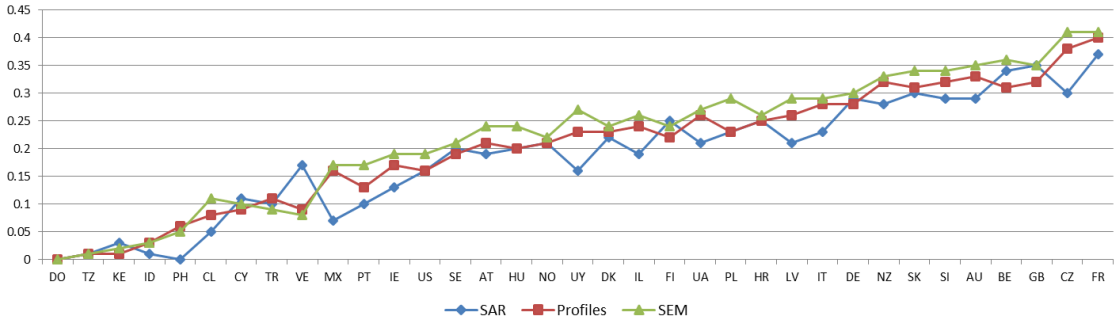

Figure 4: Adj. R^2 values for the regressions SAR/Profiles/SEM – confidence in churches and religious organizations (D)

The graphs illustrate the summed up results of Table 2 in detail: Figure 1 shows that in the majority of countries the profiles explain attitudes towards abortion a bit better than the SEM

and SAR measures, while for explaining attitudes towards homosexuality both MI measures work equally well and, generally, they work better than the SI measure SAR.

In Figures 3 and 4 the SEM line is clearly above the profiles line and decidedly above the SAR line, which means Items C and D are best explained by the SEM factor score in the vast majority of countries. Exceptions mostly take place in countries where none of the measures explain much anyway. In countries with average R^2 values greater than .2 exceptions are rare. Noteworthy are the results for Israel, where the explained variance for items A and B are on the same high level for all three measures. Also, Great Britain and Flanders stand out with particularly high levels of variance explained by SAR for item D.

Beyond this, we can see that for items A, B and C generally the more religion matters for explaining the dependent variable, the larger the gap between SI and MI measures. Item D is the exception whereby this trend cannot be observed. Nevertheless, despite that general pattern, it should be noticed that especially for item A there are a number of cases where the predictive value for the SAR indicator is at least as high as one of the MI measures.

In Indonesia, the Philippines, Tanzania, the Dominican Republic, Kenya, and Venezuela none of the three measures for religiosity explain really anything at all.

4.6 Discussion

Although, the predictive potential of the SAR indicator turned out to be better than expected, it is generally lower than the MI measures, which does not come as a big surprise. The questions that arise are: 1.Why does SAR work better in some countries than in others? 2. Why does the SEM factor explain the external factors better in some countries and in others the profiles? Is there any pattern underlying these results? And last but not least, 3. Why are attitudes towards abortion generally better explained by the profiles and the other external factors by the SEM factor?

Let us first turn our attention towards those countries where we see just about no, or, at most, very weak correlations across all measures: the Philippines, Indonesia, the Dominican Republic, Tanzania and Kenya. Here the correlations do not give us any reason to believe in great reliability of any of the religiosity measures. Additionally, in these countries all three measures for religiosity explain almost nothing about the external factors. However, the reason the results appear that way is probably not that religiosity cannot be captured with the measurement instruments or that religiosity does not play a role for attitude development in these countries. On the contrary, the data shows that people in all these countries are

generally very religious. That, however, leads to very little variance in the single items measuring religiosity. The four countries are among those with the highest SAR mean value and the lowest Standard Deviation (see Appendix). Without variance, there are no slope parameters which could be explained by regressions. Therefore, for highly religious countries these kinds of analyses, unfortunately, do not provide any meaningful results and testing hypotheses on their base should probably be avoided.

The majority of countries, however, show substantial correlations between the measures, which speaks for their reliability, and substantial explained variance for the external factors. It is easy to imagine various country-level factors that might structure patterns of each measure's performance. The most obvious factor is probably the countries' denominational structure. While regional, historical or cultural aspects shared by some countries and not by others, could also be assumed to play a role. However, I cannot recognize any such pattern. There is no systematic difference between predominantly catholic or protestant countries. And there are no regional lines that could be drawn or historically or culturally close countries that could be clustered. The crucial element of which indicator works best seems to be the content, in other words the dependent variable.

For example, in the most secular country in the sample, the Czech Republic, all three measures of religiosity explain much more variance of the directly religion related items (C and D) and not so much with regard to attitudes towards abortion (A) or homosexuality (B). It seems that Czechs, even if they consider themselves to be religious, only feel their religiosity actually addressed when it comes directly to religious matters. They do not necessarily take on the official church endorsed viewpoint towards abortion or homosexuality; however that is not a general pattern for secular countries.

Loo (2002), Bergkvist and Rossiter (2007), Diamantopouloset al. (2012), and Postmes, Haslam and Jans (2013) argue that SI measures are particularly suitable when the construct to be measured is quite homogeneous and not too complex. One might assume now that in some countries, especially in mono-denominational countries, religiosity might be more homogeneous than in others. Consequently in these countries the SAR indicator should generally work better. But the results do not support that. In Israel (85% Jews) SAR works particularly well explaining attitudes towards items A and B. For the other two items it is less potent than the MI measures. In Sweden (66% at [least nominal] Protestants) SAR works very well for explaining A and attitudes toward trust in science vs. faith (C), while in the religiously highly pluralistic US it works very well for A and rather insufficient for B and C. In Venezuela (79% Roman Catholics) there is no predictive potential of any of the three religiosity measures

for the items A, B and C. All of the measures explain some variance just for item D, which again speaks for the fact that it is the content of the dependent variable that makes the difference on a national basis.

If SAR differs a lot from the two MI measures we can assume that this depends on the component(s) the majority of respondents in the countries refer to when answering the SAR question. An example for this could be Uruguay, which we know has a decent population (23.2%) reporting belief in God, but not declaring having a religion (Enstituto Nacional de Estadística Uruguay 2006). That means it can be assumed that when asked how religious they are, a comparably high percentage of people refer to their belief in God, leaving aside church institutional aspects. This could lead to the comparably high explanatory potential of the SAR indicator on item C asking for religious faith and the observed opposite effect for item D related to trust in institutions. For the query investigating confidence in churches and religious organizations, the MI measures that include the institutional aspects have the far better predictive potential. Unfortunately, we never know for sure which religious aspects the SAR indicator contains on the individual level and consequently, also do not know which components are most important on the country level. Interpretations will therefore always be speculative. To solve that problem it might be worthwhile, and definitely highly interesting, to conduct cognitive interviews across countries, asking people exactly what their answers to the SAR question are based on. This qualitatively collected information would increase the chances for valid interpretations, but of course, leads us back to the dilemma of collecting information vs. spending financial recourses.

Also for the question of why in some countries the SEM factor explains the external factors better, and in others the profiles, there are no systematic patterns for an answer. If the profiles reach higher predictive values for items B and C than the SEM factor, the difference is generally very small. The only real difference, again, is content related. Profiles predominantly explain attitudes towards abortion better than the SEM factor. Thus, it could be speculated that there is a specific profile across countries shared by supporters of women's rights, which is, at least in some countries clearly differentiated from LGBT rights. Otherwise we would see the same patterns for attitudes towards homosexuality.

Australia is the only country which offers a striking exception from these general patterns. Looking back at the results of the variance analyses, Australia is the only country where highly non-significant differences between two profiles with respect to the mean values of both continuous measures can be observed. Australia is also the only country where for all but one indicator (D), profiles work better than the SEM factor, while the SAR indicator works

particularly insufficiently. The two profiles have very similar mean values for the continuous indicators, but one profile is indeed very special. Respondents showing this religious profile could be called 'church-free spirituals'. Their medium SAR value is the lowest after the non-religious group as 91% report not being religious, but spiritual, and 85% report not believing in God but in a 'Higher Power'. They are as liberal towards abortion and homosexuality as the non-religious group. However, these church-free spirituals only make up 6% of the whole Australian sample. Finding out about their actual impact should be the subject of further research. Like conducting cognitive interviews, looking at the single religious profiles in detail might add very useful qualitative information to the greater picture of religious landscapes.

4.7 Conclusion

The goal for international surveys must be to find the most economic and still reliable measures for the research subjects. For this purpose, this study evaluated the reliability and the predictive potential of the SI measure *self-assessed level of religiosity* (SAR) in comparison with two MI measures of religiosity.

The first hypothesis expecting the MI measures to have higher predictive potential than the SI SAR measure could be confirmed. Only in very few cases the SAR measure performs better than one of the MI measures. How well suited SAR is as a substitute for MI measures of religiosity differs across countries and the results confirm the concerns of Diamantopoulos et al. (2012: 446) that an SI measure which may perform as well as the MI measure on one context may not do so in another. For some countries it is hard to say how well SAR captures religiosity, because even the MI measures do not work as expected. In other countries, first of all Israel, the predictive potential of all three measures for all external factors is almost on the same level. SAR would be a very suitable substitute here for long scale measurements. Mexico is the opposite example; here SAR falls far behind. The other countries lie somewhere between these poles. It is not easy to decide where to draw the line for responsibly exclusively relying on the SAR indicator. There is probably no other single indicator that works better, since the results of the paper showed clearly that what measure works best depends a great deal on the dependent variable. Though generally speaking, with a few exceptions discussed above, the SAR indicator turned out to achieve good reliability and reaches a very substantial predictive potential compared with comprehensive MI measures.

The second comparison this study looked at was between the two MI measures. Contradicting my second hypothesis, conducting LCA profiles did not significantly increase the predictive potential for the four dependent variables. For the examples in this study it is true that the

categorical measure, which is more complicated to conduct, more time-consuming in interpretation and especially far more complicated in its further implementation in analyses, does not achieve a more accurate measurement of religiosity. However, with a view to the good results of religious profiles explaining attitudes towards abortion, depending on the actual research subject it still might be worth the effort.

It is not possible to actually calculate the financial advantage of using one item versus seven, since field institutes usually offer total prices based on interview-length and the desired sample size. Nevertheless, six items cost interview time and processing effort. Thus, they make a considerable financial difference, but, as this study has demonstrated, in the majority of countries do not offer substantially concomitant increase to the predictive potential.

Comprehensive international comparability is the one aspect consciously neglected via the exclusion of Asian religions in this study. The items in the ISSP Religion survey are based on Glock's scheme of religious denominations, which has a Christian bias and therefore contradicts his theoretical claim of universality (Huber and Huber, 2012). The consequent next step of research must therefore be to find a way for evaluating the explanatory potential of the SAR indicator to a wider scope of religions.

4.8. References

4.8.1 Data

ISSP Research Group. 2012. *International Social Survey Programme: Religion III - ISSP 2008*. GESIS Data Archive, Cologne. ZA4950 Data file Version 2.2.0,doi:10.4232/1.11334.

ISSP Research Group, Muhammad Saflianto, Paul Omondi, Joseph Thavaraja, and Evangeline Wanyama. 2013. *Religion Around the World Study of the 2008 International Social Survey Programme (ISSP)*. GESIS Data Archive, Cologne. ZA5690 Data file Version 1.0.1, doi:10.4232/1.11762.

Enstituto Nacional de Estadística Uruguay. 2006*. Encuesta Nacional de Hogares Amplidada. Accessed 23 May 2017* https://web.archive.org/web/20130927091848/http://www.ine.gub.uy/enha2006/flash/Flash%206_Religion.pdf .

4.8.2 Literature

Bacher, J and Vermunt JK. (2010) Analyse latenter Klassen. In: Wolf C and Best H (eds) *Handbuch der sozialwissenschaftlichen Datenanalyse.* Wiesbaden: VS Verlag für Sozialwissenschaften, 553-574.

Bechert I. (2013) Are Religious People Happier than Non-Religious People? In: Bechert I and Quandt M (eds) *ISSP Data Report - Religious Attitudes and Religious Change.* Cologne: GESIS, 59-71.

Beckford JA. (2003) *Social theory and religion*: Cambridge University Press.

Bellah RN, Madsen R, Sullivan WM, et al. (1985) Habits of the heart: Individualism and commitment in public life. Berkeley: University of California Press.

Bergkvist L and Rossiter JR. (2007) The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research* 44: 175-184.

Casanova J. (1994) *Public Religions in the Modern World*: University of Chicago Press.

Chaves M. (2010) Rain Dances in the Dry Season: Overcoming the Religious Congruence Fallacy. *Journal for the Scientific Study of Religion* 49: 1-14.

Cornwall M, Albrecht SL, Cunningham PH, et al. (1986) The Dimensions of Religiosity: A Conceptual Model with an Empirical-Test. *Review of Religious Research* 27: 226-244.

Costa PT and McCrae RR. (1992) *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI),* Odessa: Psychological Assessment Resources.

Danièle H-L. (1994) Davie (Grace) Religion in Britain since 1945. Believing Without Belonging. *Archives de sciences sociales des religions* 88: 60-61.

Davie G. (1990) Believing without Belonging - Is This the Future of Religion in Britain. *Social Compass* 37: 455-469.

Davie G. (2000) *Religion in modern Europe: A memory mutates,* New York: OUP Oxford.

Diamantopoulos A, Sarstedt M, Fuchs C, et al. (2012) Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science* 40: 434-449.

Glock CY. (1962) On the Study of Religious Commitment. *Religious Education* 57: 98-110.

Gosling SD, Rentfrow PJ and Swann WB. (2003) A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* 37: 504-528.

Hoeppner BB, Kelly JF, Urbanoski KA, et al. (2011) Comparative utility of a single-item versus multiple-item measure of self-efficacy in predicting relapse among young adults. *Journal of Substance Abuse Treatment* 41: 305-312.

Houtman D and Aupers S. (2007) The spiritual turn and the decline of tradition: The spread of post-christian spirituality in 14 western countries, 1981-2000. *Journal for the Scientific Study of Religion* 46: 305-320.

Huber S. (2003) Zentralität und Inhalt. Ein neues multidimensionales Messmodell der Religiosität [A New Multidimensional Model of Measurement for Religiosity]. Leske+ Budrich.

Huber S and Huber OW. (2012) The Centrality of Religiosity Scale (CRS). *Religions* 3: 710-724.

Jagodzinski W and Dobbelaere K. (1995) Secularization and church religiosity. In: van Deth JW and Scarbrough E (eds) *The impact of values.* Oxford: Oxford University Press, 76-119.

Jagodzinski W and Manabe K. (2009) On the similarity of religiosity in different cultures. *International Social Survey Programme, 1984-2009: Charting the Globe*: 313-336.

Kecskes R and Wolf C. (1995) Christliche Religiosität: Dimensionen, Meßinstrumente, Ergebnisse [Christian Religiosity: Dimensions, Measuring Instruments, Results.] *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 47: 494-515.

Krause NM and Jay GM. (1994) What Do Global Self-Rated Health Items Measure. *Medical Care* 32: 930-942.

Loo R. (2002) A caveat on using single-item versus multiple-item scales. *Journal of managerial psychology* 17: 68-75.

Lutz J, Kemper C, Beierlein C, et al. (2013) Konstruktion und Validierung einer ökonomischen Skala zur relativen Messung von physischer Attraktivität (AR1) [Development and Validation of a Single-Item Scale fort he Relative Assessment of Physical Attractiveness. The Attractiveness Rating 1 (AR1).]. *Methoden, Daten, Analysen (mda)* 7: 209-232.

McGuire MB. (2008) *Lived religion: Faith and practice in everyday life,* New York: Oxford University Press.

Pearce LD, Hardie JH and Foster EM. (2013) A Person-Centered Examination of Adolescent Religiosity Using Latent Class Analysis. *Journal for the Scientific Study of Religion* 52: 57-79.

Postmes T, Haslam SA and Jans L. (2013) A single-item measure of social identification: Reliability, validity, and utility. *British Journal of Social Psychology* 52: 597-617.

Robins RW, Hendin HM and Trzesniewski KH. (2001) Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg self-esteem scale. *Personality and Social Psychology Bulletin* 27: 151-161.

Schnell R, Hill PB and Esser E. (2011) *Methoden der empirischen Sozialforschung [Methods of the Empirical Social Research],* Munich: Oldenbourg Wissenschaftsverlag.

Shukor SA and Jamal A. (2013) Developing Scales for Measuring Religiosity in the Context of Consumer Research. *Middle East Journal of Scientific Research* 13: 69-74.

Siegers P. (2012) *Alternative Spiritualitäten: neue Formen des Glaubens in Europa: eine empirische Analyse [Alternative Spiritualities: New Forms of Belief in Europe: An Empirical Analysis.]* Frankfurt/New York: Campus Verlag.

Storm I. (2009) Halfway to Heaven: Four Types of Fuzzy Fidelity in Europe. *Journal for the Scientific Study of Religion* 48: 702-718.

Vezzoni C and Biolcati-Rinaldi F. (2015) Church Attendance and Religious Change in Italy, 1968-2010: A Multilevel Analysis of Pooled Datasets. *Journal for the Scientific Study of Religion* 54: 100-118.

Wilkes RE, Burnett JJ and Howell RD. (1986) On the meaning and measurement of religiosity in consumer research. *Journal of the Academy of Marketing Science* 14: 47-56.

Wolf C. (2005) Measuring Religious Affiliation and Religiosity in Europe. In: Hoffmeyer-Zlotnik JHP and Harkness J (eds) *Methodological Aspects in Cross-National Research.* Mannheim: ZUMA-Nachrichten Spezial.

Woods SA and Hampson SE. (2005) Measuring the Big Five with single items using a bipolar response scale. *European Journal of Personality* 19: 373-390.

Appendix

Table A1: List of countries, original sample sizes, sample sizes in the LCAs, and SAR mean values

Country	N original sample	N cases included in national LCA	SAR*	
			Mean	SD
Australia	1682	1159	2.64	1.62
Austria	1009	681	3.13	1.61
Belgium (Flanders)	1246	895	2.88	1.56
Chile	1502	1294	3.40	1.33
Croatia	1198	975	3.99	1.25
Cyprus	1000	754	4.11	1.09
Czech Republic	1508	1241	2.15	1.44
Denmark	1975	1321	2,50	1.38
Dominican Republic	2084	1931	3.68	1.28
Finland	1124	728	2.80	1.55
France	2438	1625	2.36	1.53
Germany	1697	1131	2.45	1.85
Great Britain	1945	1288	2.58	1.61
Hungary	1009	834	2.84	1.34
Indonesia	1948	1825	3.85	1.08
Ireland	2046	1688	3.70	1.08
Israel	1172	879	3.13	1.64
Italy	1077	914	3.43	1.47
Kenya	1483	1379	4.60	1.05
Latvia	1069	888	2.85	1.34
Mexico	1470	1083	3.69	1.18
New Zealand	981	701	2.73	1.56
Norway	1071	652	2.56	1.47
Philippines	1116	938	4.17	0.91
Poland	1263	948	3.82	1.07
Portugal	999	851	3.67	1.31
Slovak Republic	1137	898	3.51	1.55
Slovenia	1063	800	2.88	1.61
Sweden	1230	797	2.42	1.38
Tanzania	1516	1399	3.78	1.23
Turkey	1453	1309	4.51	1.14
Ukraine	2030	1146	3.48	1.31
United States	1338	1201	3.80	1.42
Uruguay	999	920	2.56	1.59
Venezuela	1063	920	3.78	1.18

*SAR mean values on the basis of N included in LCAs (column 2)